REPORT RESUMES

ED 003 627 56

INVESTIGATIONS OF THINKING VIA SELF-INSTRUCTIONAL PROGRAMS.
BY- BRIGGS, LESLIE J. AND OTHERS
AMERICAN INST. FOR RESEARCH IN BEHAVIORAL SCIENCES
REPORT NUMBER NDEA-VIIA-1002 PUB DATE JUN 64
REPORT NUMBER AIR-D41-6-64-FR
GRANT OEG-7-48-7670-196
EDRS PRICE MF-$0.18 HC-$3.32 83P.

DESCRIPTORS- *AUTOINSTRUCTIONAL AIDS, *AUTOINSTRUCTIONAL METHODS, *AUTOINSTRUCTIONAL PROGRAMS, *METHODOLOGY, PROBLEM SOLVING, *THOUGHT PROCESSES, CREATIVE THINKING, PROGRAM EVALUATION, *PROGRAMED INSTRUCTION, *PROGRAMED TEXTS, PALO ALTO, CALIFORNIA

OBJECTIVES OF THIS STUDY WERE (1) TO STIMULATE THINKING ON THE PART OF STUDENTS BY EMPLOYING NEW FORMS OF AUTOINSTRUCTIONAL SEQUENCES AND (2) TO SEEK INCREASED UNDERSTANDING OF THOUGHT PROCESSES. IN PHASE ONE LINEAR PROGRAMS WERE ANALYZED TO IDENTIFY CHARACTERISTICS OF THOSE PROGRAMS THAT INCREASED ACTIVE RESPONSE BY THE STUDENT. THE ANALYSES SUGGESTED THERE SHOULD BE VARIATIONS IN STYLE OF SEQUENCES WITHIN PROGRAMS TO (1) DIFFERENTIALLY ESTABLISH NEW LEARNING, (2) PROVIDE REVIEW TO ENHANCE RETENTION, AND (3) GIVE MORE PRACTICE TO THE STUDENT IN ORGANIZING AND CONSOLIDATING NEW LEARNINGS. FROM THE RATIONALE DEVELOPED SOME NEW PROGRAMING FORMATS WERE DEVISED. IN PHASE TWO, SOME OF THE NEW FORMATS IN WORKBOOK FORM WERE COMPARED IN CLASSROOM SITUATIONS. IN PHASE THREE, CLOSER OBSERVATION AND MORE CONTROL OF CONDITIONS WAS ACHIEVED BY WORKING WITH INDIVIDUAL STUDENTS USING TEACHING MACHINES. RESULTS SHOWED THAT SUPPLEMENTARY PRACTICE STUDY PRODUCED NO BETTER RESULTS ON A CRITERION TEST THAN ADDITIONAL STUDY OF THE BASIC LESSON. PROBLEM-SOLVING PERFORMANCE APPEARED MORE CLOSELY RELATED TO TIME SPENT IN STUDY THAN TO VARIATIONS IN HOW THE STUDY WAS DIRECTED. KNOWLEDGE OF PRINCIPLES COUPLED WITH THE STUDENT'S OWN ABILITY APPEARED TO BE THE TWO MAIN FACTORS IN SUCCESSFUL PROBLEM SOLUTION. (AL)

AIR-D41-6/64-FR

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

1002

1002

INVESTIGATIONS OF THINKING
via
SELF-INSTRUCTIONAL PROGRAMS

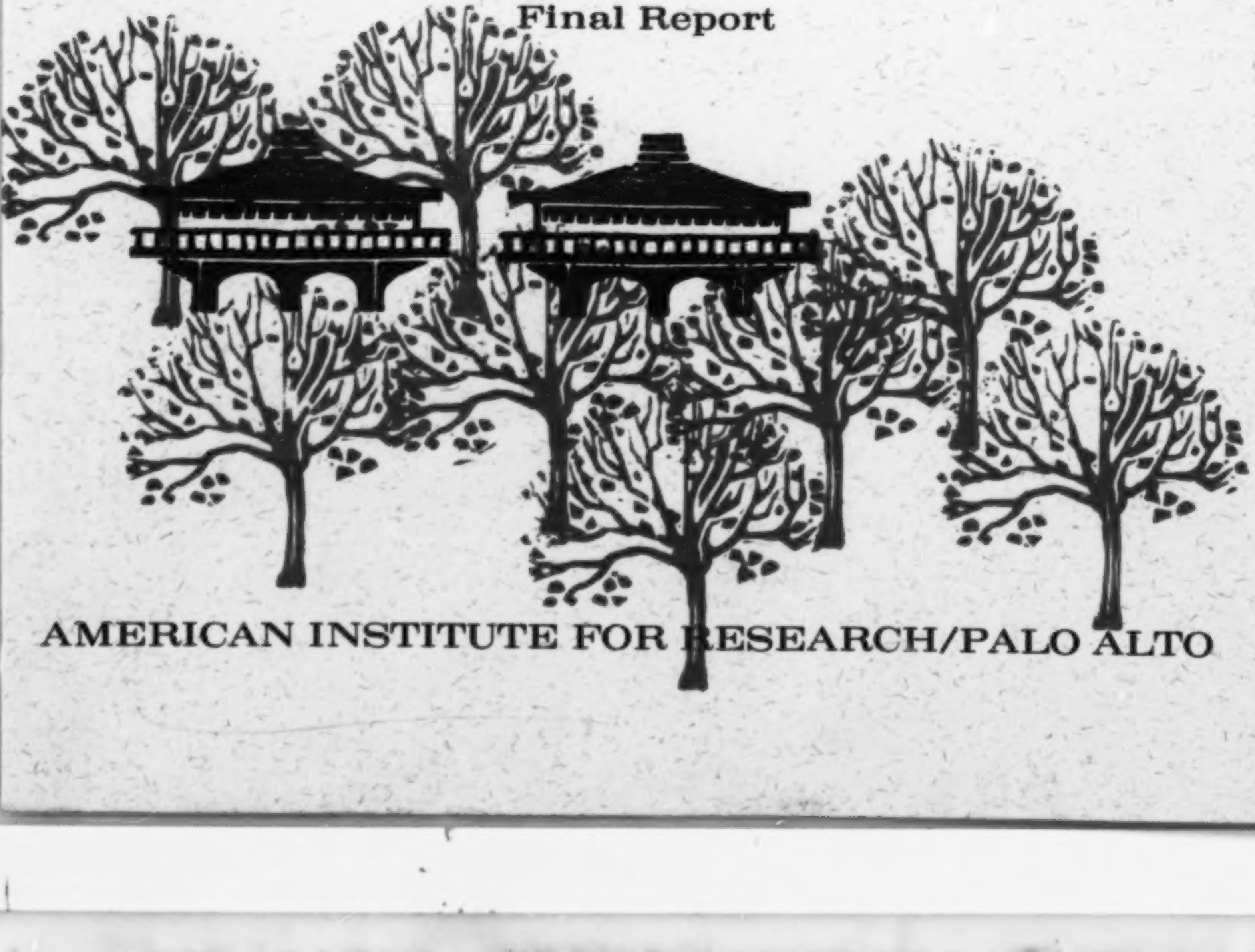
00 3627

Leslie J. Briggs
Vincent N. Campbell
Lloyd O. Brooks

June 1964

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office Of Education
Grant Number Title VII: 7-48-7670-196

Final Report



AMERICAN INSTITUTE FOR RESEARCH/PALO ALTO

0 003627 new forms of autoinstructional P. M.

~~Objectives~~ techniques were used
~~The objectives of the proposed~~ [illegible] (1) to
stimulate thinking on the part of students ~~by~~
~~employing new forms of autoinstructional~~
~~techniques~~ and (2) to seek increased
understanding of thought processes. To this
end two kinds of experimental investigations were
devised --(1) ~~the first~~ was the use of materials
and class procedures which depart ~~drastically~~
from standard autoinstructional procedures
and (2) ~~secondly~~ intensive examination of the
specific features of linear programs. In a classroom
experiment a new format was administered
under various procedural conditions. Analysis
showed no significant difference due to either
the new format or procedural technique.
Analysis of data from individual students
aking linear programs indicated that use of
esponse latency measures in conjunction with
rogram errors would promise a better basis

for program revision than the present bases of error data alone. Included in the ~~[illegible]~~ report were other completed formats which were suggested as possible bases for future studies in autoinstructional techniques.

607 Palo Alto, Calif.

* Autoinstructional Aids
* Autoinstructional Methods
* Autoinstructional Programs
* Methodology
Problem Solving
* Thought Processes
Creative Thinking
PROGRAM EVALUATION

AMERICAN INSTITUTE FOR RESEARCH
mailing address: Station A, P.O. Box 11487, Palo Alto, Calif.
Telephone 321-2130 (Area Code 415)
office location: 1791 Arastradero Road, Palo Alto, California

INVESTIGATIONS OF THINKING
via
SELF-INSTRUCTIONAL PROGRAMS

Leslie J. Briggs
Vincent N. Campbell
Lloyd O. Brooks

AIR-D41-6/64-FR

Grant Number Title VII: 7-48-7670-196
National Defense Education Act of 1958

Final Report

Leslie J. Briggs
Principal Investigator
American Institute for Research

June 1964

The Research Reported Herein was
Supported by a Grant from the
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

Foreword

The purpose of the project was to stimulate increased thinking activities on the part of the student by developing new procedures in the application of programed instruction. The work on this project was conducted by a sequence of three distinctly different ways of exploring issues relating to thinking and programed instruction; it therefore appeared desirable to prepare three separate chapters, arranged in the order in which the work phases actually occurred.

The first, exploratory phase consisted largely of devising the rationale for new kinds of programing formats and strategies which were intended to stimulate increased active responding by the student. The purpose was to supplement normal programed sequences with exercises which require active recall and organizing efforts on the part of the learner. The first chapter presents samples of these program formats and the rationales upon which they were based. Much of the work on program formats reported in Chapter I was performed by Arthur W. Krauser and Mary B. Willis, with assistance from Zita Brown.

The second phase of work utilized some of the program formats and materials, developed in the first phase, in a classroom experiment designed to compare several alternate ways of stimulating problem-solving activities by the student. The suggestion for this kind of experiment was made by Mary B. Willis, who also prepared much of the materials and tests employed in the experiment. The experiment was designed and conducted by Vincent N. Campbell, the author of Chapter II, with assistance in data analysis by Donald F. Terry.

The third phase of research brought to bear the closer observation and control of conditions made possible by working with individual students in a well-instrumented environment. The work of Lloyd O. Brooks during the third and final phase of the project is presented by him in Chapter III.

While it cannot accurately be said that these three distinct phases of work were foreseen at the outset of the project, in the opinion of the writer all three have been valuable. Phase I was conducted to explore rationales and to work with program materials without any restricted theoretical or program-style boundaries. Phase II of the work took advantage of the economics of gathering data in groups under realistic classroom conditions, under the control of the regular teachers in the participating public schools. Thus Phases II and III were separate efforts to capitalize upon the special advantages of conducting research with groups and with individuals.

Due to the complexity and magnitude of the problem, the total project was expected to be exploratory rather than definitive in nature. It has been possible, therefore, to derive only some threads of theory which can be traced through all phases of the work. We trust that the sample program formats as well as the experimental results will interest other investigators in further study of some aspects of the dual problem of increasing the thinking activity of students through use of programed instruction, and the use of programed instruction as a vehicle for the further study of complex thought processes.

Finally, the writer expresses his appreciation to all the participants in the work of the project. It is perhaps remarkable that the variety of approaches and points of view utilized by the project members during the three phases of work, while reflecting a general characteristic of the present stage of research knowledge, did not detract from close cooperation in the search for further understanding of the conditions which enhance the kind of learning under investigation.

Leslie J. Briggs
Principal Investigator

TABLE OF CONTENTS

Page

FOREWORD i

Chapter I: Problem, Materials, and Rationale
by Leslie J. Briggs 1

Chapter II: Learning to Solve Problems:
A Classroom Experiment
by Vincent N. Campbell 12

Appendices to Chapter II 18

Chapter III: Laboratory Study of Individual
Performances on Programed Learning
Tasks
by Lloyd O. Brooks 32

Abstract 70

References 72

Chapter I

Problem, Materials, and Rationale

Leslie J. Briggs

The Problem

The overall objectives of the project have been (a) to stimulate thinking on the part of students by employing new forms of autoinstructional sequences, and (b) to seek increased understanding of thought processes through experimentation with autoinstruction techniques.

This chapter sets forth some of the lines of thought which led to the establishment of the project and guided the course of its development. As will be seen, the results of the general rationale led to two kinds of experimental investigation: (a) use of materials and class procedures which depart drastically from standard autoinstructional procedures (Chapter II), and (b) intensive examination of the specific features of linear programs which may be regarded as descriptive dimensions of active responding (Chapter III).

Rationale for Format Innovations

While proponents of the now-conventional linear programs have presented a theoretical defense of their methods (in terms of overt responding, self-pacing, prompting to insure successful responding, and reinforcement), several lines of empirical data appear to cast doubt on the generality of many points of the theory. Thus, overt responding has often been found no more effective than covert responding or reading, random sequencing of frames has sometimes been found equal in effectiveness to the programer's "logical" sequencing, and removal of the feedback term often does not appear to reduce learning.

It is clear that the contents of programs in such research, and the actual forces at work in the various programs, are much more heterogeneous than may be suggested by the fact that all are linear programs. Put differently, one can program in such a way as to make feedback important or unimportant (Hamilton, 1964), to make overt responding superior or inferior to covert responding (Goldbeck and Campbell, 1962), to make a random sequence as effective as a "logical sequence" (Levin and Baker, 1963), or to make moderately difficult frames more effective than easy frames (Goldbeck and Campbell, 1962).

These findings are cited to illustrate several points: (a) the search for "always-true" findings about learning is probably fruitless; (b) the content and total programing technique interact with single-variables in the total technique; (c) different programs involve different kinds of learning which interact with program style; (d) it will continue to be difficult to reach generalizable or even describable conclusions until there is developed a taxonomy of learning objectives so that the characteristics of the learning to be achieved can be identified and matched with required learning conditions (Melton, 1964; Gagne, 1964).

In the absence of an accepted taxonomy, for illustration here we may consider three kinds of learning: (a) rote memorization, (b) learning of a principle, and (c) using principles to solve problems. The evidence suggests that rote memorization is enhanced by direct prompting and repetition; that learning of principles is enhanced by less direct prompting and liberal use of examples; and that ability to solve problems is enhanced by some form of relatively unprompted guided discovery and by careful sequencing of subordinate knowledges (Briggs and Angell, 1964; Briggs and Hamilton, 1964). It therefore follows that no one set of conditions, as employed in any one program style, would be optimal for all three forms of learning. Considering further that a taxonomy adequate for general educational purposes might have eight or ten, rather than only three kinds of behaviors and correlated learning procedures (Gagne, 1964), the use of any single set of learning conditions, in an across-the-board fashion, appears indefensible.

It is not the purpose of this chapter to be overly critical of the theory underlying the linear or any other program style. It is the purpose to suggest (a) that various programs can differ drastically while having a similar superficial appearance, and (b) that some learning goals may best be met by gross departures from present program styles. When one considers together individual differences among learners and learning process differences associated with different criterion behaviors, it is even more reasonable to expect the optimal conditions of learning to vary from lesson to lesson and from student to student.[1]

Next we consider, one by one, some characteristics of conventional programs in order to illustrate in more detail our rationale for varying program style in accordance with task and criterion considerations.

Uniform Size of Step

While many writers have agreed that "size of step" has no uniform meaning, a characteristic of the conventional linear frame is to present only a small amount of new information along with prompts in order to elicit the desired response. While this kind of frame may be useful to introduce a new technical term or a new concept, frames with other features may be preferable for testing recall of previous information, particularly when it is desired that the student put together several elements of earlier instruction to form a structure of the elements, to organize them meaningfully, or to use several points of information as they bear upon a new problem. While conventional programs use "test frames" to be sure the student can recall information without full prompting, and "panels" to help the student understand how an element of the instruction fits into some overall outline or scheme of things, there is seldom a request for the student to produce an outline, to recall several things at a time, and so on.

---

[1] While research in programed instruction seeks to illuminate the kinds of task by learning condition interactions discussed here, there needs to be increasing attention to what needs to be taught as well as how it should be taught. The identification of subordinate knowledges needed to reach criterion performance often tends to be neglected while investigators are preoccupied with such issues as size of step, mode of response, and conditions of feedback. That attention to both factors is justified has been demonstrated (Gagne, 1962; Gagne and Brown, 1961).

A reflection upon study methods one often uses, say in conjunction with study by a textbook, will bring to mind such techniques as pausing at an end of a section to try to recall or recite several important points just covered. The absence of this kind of demand for a large-size response unit in linear programs suggested the need to intersperse such exercises as would require larger, organizing kinds of response units with the normal linear sequences. The "knowledge matrix," illustrated later in the chapter, emphasizes this departure from small-step exercises.

Uniform Difficulty

Many students appear to find the usual linear program tiresome, flat, or boring. This may be due to program style as well as to uninteresting content. A tennis game would be dull if every return of the ball required the same effort in the same part of the court. The popular saying about variety and spice of life may not be too inappropriately applied to programs. While no experiment is known which compared frames of varying difficulty with frames of uniform difficulty, our suspicion is that variety in difficulty, format, and size of response units would in itself be desirable. This appears especially likely when one expects some sequences to introduce new information, some to provide spaced practice to aid recall, and some to use past learning in a problem-solving situation.

Several writers have recommended that, in general, program frames should cue responses close to threshold (Lumsdaine, 1961; Goldbeck and Briggs, 1962). Yet the standard procedure is to revise programs to reduce difficulty by use of error data; there is no corresponding effort to revise overly easy frames to keep difficulty high enough to be close to threshold. Chapter III presents data concerning use of response latencies to improve programs in this latter respect. When frames are fully prompted there appears a tendency for the student to neglect the material in the frame other than the prompt term; consequently correct answers are not given on a criterion test even though the response itself has been learned (Briggs et al, 1962). Apparently when the frame requires a moderately difficult searching effort, some of the undesirable "response orientation," which leads the student to neglect some of the information, is overcome (Goldbeck et al, 1962).

It is suggested that use of something like the "knowledge matrix," described later, would cause the student to be oriented toward the future requirement to respond in contexts other than the linear "frames," and so to pursue his learning in a more useful fashion.

Continuity of Learning

Students often experience a feeling of being "lost" or not knowing what the overall structure of the lesson is when they are not provided some kind of outline. Some persons have been observed to produce an outline for themselves as they proceed through a linear program, in order to benefit from the orientation it brings. Thus, while students may be responding correctly, it may be doubtful whether they could, at any given point in the program, adequately summarize past material or report a sense of direction in their study. The need for pausing to review, in a larger sense than encountered in test frames in the program, then was another factor in the perceived need for exercises to provide such responding in a broader context.

Study as a Mixture of Learning and Review

In one widely used how-to-study program (Robinson, 1941), the code name "SQ3R" is used to remind the student, when studying a chapter in a textbook, to: (1) survey the chapter by noting headings which form a general outline of the chapter; (2) turn each heading into a question to give a goal to be reached while reading; (3) read one section or unit; (4) recite to himself the answer to the question posed; and (5) review the entire chapter by re-reciting the main points under each question posed.

This method of study alternates reading with recitation and review, all directed toward studying actively to learn and retain the key materials.

Few, if any, autoinstructional programs employ the entire strategy of the SQ3R method of study. No known experiment compares that method of study with learning from an autoinstructional program. While there are a few features in common between autoinstructional programs and the SQ3R method, the differences appear greater than the similarities. Actually, the variations from linear programs presented in this chapter constitute, generally, a step toward integration of programing with effective study methods.

While both the branching and the adjunct forms of programed instruction appear closer to the SQ3R method than does the linear form, surprisingly little research has been done in recent years with those as compared to research with linear programs.

One exception to the above is a study by Hershberger and Terry (1964) who do emphasize that study is really a mixed process of learning and review. They make the hypothesis that the manner in which self-test items are employed in programs should differ in accordance with the intent to use them for clearing up misunderstanding or to provide practice to enhance recall.

In two recent reviews of the experimental literature dealing with programed instruction (Briggs and Angell, 1964; Briggs and Hamilton, 1964), a number of instances were cited in which the learning conditions which provide most effective or efficient initial learning are different from the conditions of practice which enhance retention. It appears generally that learning research has failed to utilize retention or transfer as criteria as often as the facts would merit. As a separate matter, it appears likely that the overall scheduling of review following programed instruction might differ from optimal scheduling of review following learning by conventional methods, as revealed by earlier research. One project now in progress is investigating this matter.

Format Innovations: Sample Materials

To summarize previous introductory remarks, the first phase of the project consisted of logical analysis of conventional programs with respect to two possible means for improving the thinking activities of the learner: (a) variations in specific frame characteristics, such as response latency (Chapter III), and (b) gross variations from conventional programed procedures (Chapter II).

While Chapters II and III present results of experimental work conducted along the two lines of variation from conventional programs, the remainder of this chapter will present other samples of format variations which were not used in formal experiments, but which were tried out informally with sufficient individual Ss to be able to state that they are feasible for further investigational use inasmuch as students are able to complete them successfully. However, not enough work was done with these particular formats to evaluate them further, or to state more explicitly the exact benefits they may bring when employed jointly with linear sequences.

Response-Response Sequences

The specific rationale for these materials was that the S having proceeded in stimulus-response fashion through a basic linear sequence, should be prepared next to review and summarize the material learned by receiving only a sufficient stimulus to start him to responding. The direction of relationships should at first be indicated for him, after which S should be able to form his own directions, given a rather open set of conditions. Thus, his first response, rather than a programed stimulus, provided the direction for the next response. Hence the coined term, response-response program.

The simplest form of this program style was presented to S as a game involving words and arrows. In the sample below, the S had first completed a linear sequence dealing with temperatures of water, forms of water, molecular processes associated with these, and rain cycles. In this first sample, an S would first be supplied the initial terms ("liquid water"), and either one of the two remaining terms ("very cold" or "solid water"). S was to supply the missing term. Next, S was supplied any two of the terms, so he would have practice in supplying any one of the three terms, regardless of the direction of the arrow.

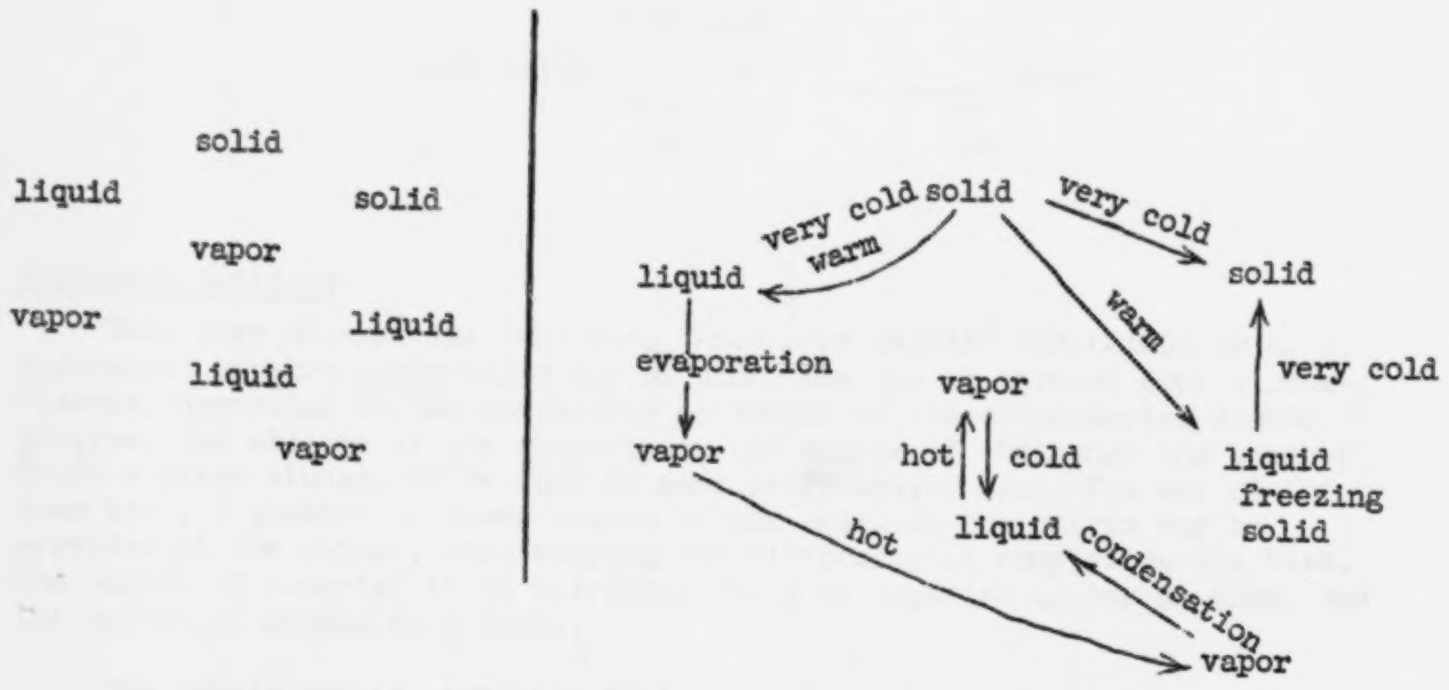
liquid water $\xrightarrow{\text{very cold}}$ solid water

solid water $\xrightarrow{\text{melting}}$ liquid water

slow molecules $\xrightarrow{\text{hot}}$ fast molecules

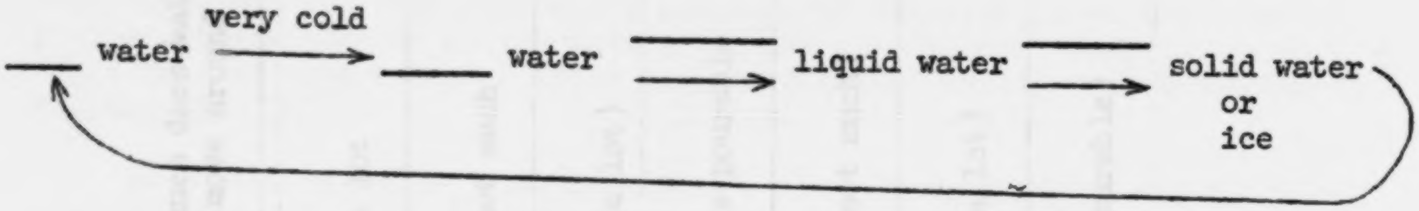
clouds $\xrightarrow{\text{condensation}}$ rain

For S̲s who needed to do so, reference could be made back to the linear sequence. For those who succeeded easily without such reference, the next, more openly structured form of the words and arrows game was presented. S̲s were told that they may draw more than one arrow from any word and that they may draw arrows back and forth between the same pair of words.

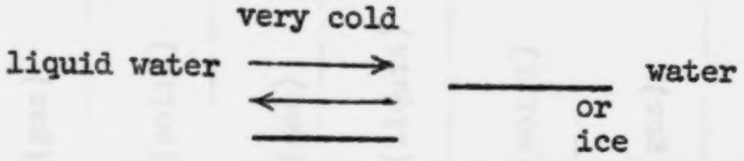

solid
liquid
solid
vapor
vapor
liquid
liquid
vapor
very cold
solid
very cold
warm
liquid
solid
warm
evaporation
vapor
very cold
vapor
hot
cold
liquid
freezing
solid
hot
liquid
condensation
vapor

Thus, the series of two forms of this exercise provided a variety of stimulating conditions and much freedom to demonstrate and practice the knowledge of relationships and processes gained from the initial instruction.

Another form asks Ss to fill in the missing words in a sequence.



Final examples of any form emphasize the reversibility of the reactions or process.



Knowledge Matrices

This type of exercise (the term "knowledge matrix" was coined by A. A. Lumsdaine) appears potentially useful under any one or more of four circumstances, depending on the complexity or length of the accompanying linear program, the ability of the student, or the degree of challenge the teacher finds a given student to be able to meet profitably. Also, for any of the four uses, a greater or fewer number of the cells in the matrix may be provided at the outset, thus varying the difficulty of completing the task, the amount of material it is desirable for S to organize as one problem, and the amount of prompting S needs.

The sample matrix, entitled "Molecules," can be employed in any of the following ways:

(a) As an adjunct (panel) to the linear program; the S refers to it in much the same fashion that an outline is used. The program and the matrix could be corss-indexed, if desired, so that when a correct response is elicited by the program, S is instructed to also write the word in the appropriate cell. This could help S see the structure of the program, and provide two contexts for practice as he goes through the program.

(b) As an independent review exercise after the program; S could be instructed to refer back to the program only when he cannot fill in a cell.

(c) As a test following completion of the study.

- Molecules -

heat	speed	move around	how spread out	water form	weight	how much does water form move around
hot	(very fast)	(a lot)	very	(gas)	(very light)	a lot
(very cold)	(slow)	not much	(close together)	(solid)	(heavy)	not much
(hot)	very fast	(a lot)	(very)	(gas)	very light	(a lot)
warm	medium	(a bit)	(loosely)	(liquid)	heavy	it's pourable
very cold	(slow)	(not much)	close together	(solid)	(heavy)	(not much)
(hot)	(very fast)	(a lot)	very	(gas)	(light)	(a lot)
medium	(medium)	(a lot)	loosely	(liquid)	(heavy)	(pourable)

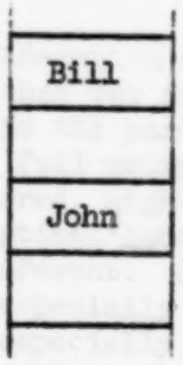
Words in parenthesis are to be filled in by Ss

(d) As a practice exercise with diminished prompts; S could at first be given only a few cells to fill in after completion of the linear program. Then a new matrix, with more cells missing, is supplied, with the eventual goal of filling in the entire matrix without assistance. Such an exercise might increase the value of later problem-solving exercises as described in Chapter II.

It should be noted that the entire matrix about molecules need not be used at once, if it is desired physically to incorporate matrix exercises in the program itself rather than using it in an adjunct or "panel" fashion. In that case, at the end of a sequence on "ice," for example, only the row in the matrix describing ice would be used. It appears likely that having learned the principles associated with ice, some students might be able to by-pass the linear sequence on "water," and to correctly reason out the proper matrix entries. Considerable adjustment could be made to individual differences in this fashion without threatening the less capable students because of too little prompting or without requiring unnecessary sequences of work for fast learners.

Logic Ladders

This term was coined to identify a type of small "panel" which was employed in conjunction with a linear program sequence. The purpose of the "ladder" is to provide a visual framework which shows quantative relationships on one dimension more clearly and in a more organized fashion than is the case if the elements are constrained into the format of conventional linear frames. Tried with a few Ss, such items appeared motivating (for young Ss at least), and appeared to aid in retention.



Example:

"Look at the ladder at the left. You can see that Bill is taller than John. Jack is shorter than Bill, but taller than John: write Jack's name in the ladder."

(Response and feedback.)

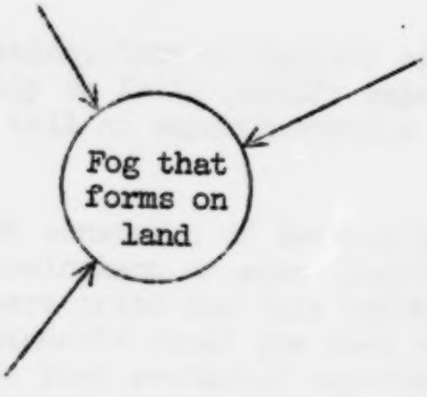
"Try to write the three names, from tallest to shortest without looking at the ladder; but if you need to look, do it."

(Response and feedback.)

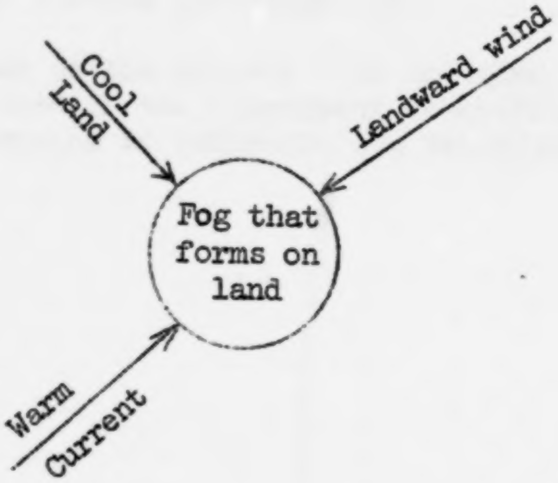
"Tom is shorter than John. How many boys are taller than Tom?" (Etc.)

Causation Diagrams

Another device we have called a causation diagram. In this the factors contributing to an end result are represented by arrows and the end result itself by a small circle. For example, a child may be given a causation diagram, such as the one below, with only the end result filled in, and asked to label the arrows. If he were given the following diagram:



he would fill it in to read:



Alternatively, he might be given a causation diagram with the causes, as represented by the arrows, filled in, but with the end result, in the circle, not indicated. He would then label the circle that stands for the end result.

Summary

A consideration of the features of conventional programs has led to the impression that two forms of departure from them might improve the thinking stimulated on the part of the student. First, departure from the usual easy, small step, full prompted frame, and from the rather invariable evenness of effort required, might better serve the dual purposes of learning and retention, the optimal conditions for which, experimentation has often shown, are somewhat different. Second, more gross departures from the usual linear format may especially be useful for stimulating thinking and enhancing retention, especially by alternating linear sequences for initial learning with more problem-oriented exercises to enhance thinking, organization, and active review and rehearsal.

Considerable variations in program style are also justified on the grounds that different tasks require different learning conditions. Particularly needed are forms which encourage study efforts similar to those taught in effective study courses--exercises which make use of the features of outlines, thought problems, and organization and review of elements of material initially learned as more discrete elements.

A third source of variation, that of ability among learners, added to the others, seems logically to fully justify experimentation with flexible linear programs as well as experimentation with new program formats and procedures.

This first phase of work consisted of developing rationales for program innovation and of development of some illustrative program formats. These illustrative formats were tried out only informally with a few students in order to determine that students could use them successfully. Thus, these sample formats have not been evaluated experimentally, but are described in this chapter to encourage further investigation.

The preliminary work of this phase of the project also included development of some of the materials used in the experiment described in Chapter II, and development of some aspects of rationale and materials employed in Chapter III.

Chapter II

Learning to Solve Problems: A Classroom Experiment

Vincent N. Campbell

The systematic nature of programing seems especially appropriate for the study and enhancement of problem solving, as opposed to divergent thinking or fluency of ideas, because both programing and problem-solving seek to converge toward fairly clearcut objectives. In the present study we compared different self-instructional programs intended to teach problem-solving in a normal classroom within an existing content area of the curriculum.

A problem is usually solved when certain ideas come to mind contiguously. Before this can happen the problem-solver must have formed a repertory of ideas which includes at least one subset of ideas sufficient for solution of the problem. (There may be more than one subset of ideas which works). The goal is to have at least one such "sufficient subset" of ideas occur within a reasonably short time after presentation of the problem.

If one were instructing an S to solve one particular problem the task would appear straightforward--activate the ideas symbolically, show him referent objects, literally tell him how to solve it--but the main educational objective is to be able to solve any of a near-infinite class of problems generated by some specified range of knowledge and situations. Concentrating initial training on underlying ideas and principles seems a worthwhile subgoal to pursue when a reasonably small set of ideas can be used to solve a much larger set of particular problems. But given that such ideas have been formed, what is the best means of causing a workable combination of ideas to occur on presentation of any one of a large class of related problems, not knowing in advance what the particular problem will be? In other words, how does one learn to solve problems? For this purpose, how valuable is practice in solving other problems from the same class? Would such practice time be better spent assuring more thorough learning of the ideas and principles upon which most solutions are based? This was one of the questions posed by the present experiment.

The other experimental questions concerned how one should practice solving problems in case practice in problem-solving should prove superior to concentrating on principles (combinations of ideas). Practice could concentrate on one or more of the many subprocesses which have been suggested as leading to problem-solving, e.g., orientation, preparation, analysis, incubation, hypothesis evaluation, searching and question-asking, set-changing, and guidance. Russell (1960) has reviewed studies of problem-solving and concluded that training in problem-solving does not generalize broadly; rather it must be oriented toward a specific process or type of problem if it is to be effective. More recent evidence (Gagne and Dick, 1962; Kersh and Wittrock, 1962; Anderson and Anderson, 1963) tends to confirm this, except for the findings of Crutchfield and Covington (1963).

Their programed course on solving detective-type problems emphasized development of question-asking skills, and it seemed to significantly facilitate the solving of test problems in quite different content areas. In other recent studies of problem solving the practice variable most studied perhaps (and with conflicting results) has been the degree of "guidance" in discovery of problem solutions.

Degree of guidance (or hints) and question-asking were therefore chosen for experimental treatment in the present study. These two variables were combined factorially yielding four different conditions of practice in problem-solving: problems alone, problems with hints, problems with question-asking, and problems with both hints and question-asking. Two other groups were not given practice problems, but instead thoroughly reviewed ideas and principles, one group being encouraged to ask questions and the other group not. The six different learning conditions encompassed in the total design are diagrmed below:

	Question-asking	Work alone
Problem + hints	Group PQH	Group PWH
Problems alone	Group PQ	Group PW
Review of principles (no application problems)	Group KQ	Group KW

Subjects

Seventy 4th grade students who were being taught geography by the same team of teachers served as subjects (Ss). The experimental unit of instruction was part of their regular course in principles of global geography. The students were heterogeneous in tested ability (CTMM , CAT) although the overall means on these tests were somewhat above the national norm.

Method

The general educational objectives of the experimental unit were to be able to solve any of a large class of applied problems concerning the climate and conditions of life in the horse latitudes, mainly as compared with the doldrums region, and to know the core ideas and principles which form the basis for solving such problems.

Uniform initial instruction. It was assumed that some initial instruction intended to impart the main ideas would be more effective than starting Ss immediately on difficult application problems. Therefore, all Ss were given a regular programed instruction sequence on the horse latitudes for the first 3 days of instruction (45 minutes per day). Two illustrative pages of this initial instruction are shown in Appendix A. The Ss were quite familiar with the use of this program format from previous geography units. The program required frequent responding and emphasized prediction rather than verbal definitions. The answers to each program item were covered by S's answer sheet, which he slid downward to reveal each correct answer after his own response. Ss were seated in groups of 6 (for initial instruction only) and could seek help from other Ss in their own group when having difficulty.

Differential learning treatments. After the initial programed instruction, the 70 Ss were divided randomly into 6 groups corresponding to the 6 cells of the diagram on page 13. All Ss were told that the purpose of the next 3 days of instruction was to get them to "think about and be able to use" what they had just learned about the horse latitudes. Study time (three 45-minute periods) was equated for all groups, rather than number of problems or questions, because previous research (e.g., Campbell, 1963) had often shown that study time is the variable most predictive of amount learned by different methods.

The total study space was then separated into 2 classrooms by a movable partition. One room became the question-asking room (treatments PQH, PQ, and KQ in the diagram); the other room became the work-alone room, in which questions were answered when asked (very rarely as it turned out) but were not encouraged.

Ss in the question-asking room were told that learning to ask questions about a problem, as a way of trying to solve it, was to be their main objective. Ss in problem-solving groups PQH and PQ were required to ask at least 3 questions about each problem before writing their answer to the problem. Ss in the KQ group, who were given thoughtful review questions rather than application problems, were encouraged to ask questions but not required to ask any. From 3 to 5 experimenters and teachers were circulating about the room at all times privately answering an S's questions as briefly as possible or referring S to his program or atlas. Questions were not criticized nor reacted to negatively, and we tried to reward by private praise any question showing any merit whatsoever. Ss could not be seated far enough apart to ensure that question-asking and answering would not distract other Ss nearby. The general masking noise of several simultaneous interactions may have helped to prevent such distraction. Whether the general buzz itself impaired learning is hard to say.

In both rooms a question asked by S was indicated by E with a checkmark in colored pencil on the program. Ss were told to work carefully and that they did not have to finish any given amount of material.

The Knowledge-Review groups (KQ and KW) were given the "matching game" and "tell why" exercises shown in Appendix B. After each exercise was completed it was checked by a teacher without correction and S was required to restudy certain parts of his basic program and atlas until all items in the exercise were answered correctly. It was intended that these exercises be thought-provoking and stimulate approximately as much intrinsic interest as the application problems did in the other groups. There seemed already to be sufficient evidence (e.g., Haslerud and Meyers, 1958) that rote review of principles would be less effective than a more varied approach.

The practice-problem training groups (PQ, PQH, PW, PWH) were all given a program in which the same practice problems (28 in all) were arranged in the same order. The order of the problems was essentially random, so that Ss completing only part of the sequence would have had problems drawing on all the main ideas of the unit. Each problem appeared on the front page of a $5\frac{1}{2}$" x $8\frac{1}{2}$" booklet. The remaining pages contained hints (one per page) for the hint groups (PQH and PWH) but were blank scratch paper for the no-hint groups (PQ and PW). For the question-askers (PQH) each hint appeared in the form of a hint-question, which any E would answer at S's request. In the work-alone room the same information was contained in each hint in the form of a declarative statement. For the hint groups each problem booklet was stapled in the middle so that each page had to be torn loose in order to read the next hint page. This provided a count of the number of hints used by each S on each problem. There were an average of 3.7 hints available per problem and both the PQH and PWH groups actually used an average of about 1 hint per problem.

The instructions for each problem group are shown in Appendix C. Following each instruction page, pages 1 and 2 of Problem Booklet 1 are shown for the corresponding learning condition. The first ten practice problems and their hints are shown in condensed format in Appendix D.

On the third and last day of differential training a few Ss from each group finished the last of the training materials before the end of the period, but the differences between groups in number of Ss finishing early was not judged to represent an appreciable departure from the premise of equal average training time for all groups.

Four days after the end of training all Ss were given the same 4-part criterion test. Part IV consisted of a brief multiple-choice test over the main principles and facts covered by the horse latitudes unit. The first three parts of the test consisted of application problems drawn separately from the same population as were the training problems. That is, in planning the experiment, a pool of items constructed for each major content area of the unit was divided randomly into two groups, one arbitrarily designated to be training problems, the other criterion test problems.

In Part I of the test, Ss were given 9 application problems to work without any hints or outside aid. Part II consisted of 7 problems, each one followed by a series of rather direct hints, that is, statements of relevant information analogous to those given Group PWH during training. In Part III Ss were given 6 problems without hints, but all Ss were permitted and encouraged to ask any E as many questions as they wished in trying to answer the problems. Part III thus corresponded to training condition PQ, except that Ss were not required to ask any certain number of questions about the test problems.

Part II of the test was intended to assess whether the presence of hints during training had a more beneficial effect on subsequent problem-solving when hints were retained as part of the criterion test. Likewise, Part III was intended to assess whether question-asking during training would be more effective for criterion problems when the question-asking privilege was retained.

All Ss began each part of the test simultaneously. Total administration time was 70 minutes.

Results

All test protocols were scored by two different scorers without knowledge of the learning conditions to which the Ss belonged. On every problem a maximum of 3 points was given. Most problems asked the S to tell why his answer was correct. In such cases S was given one point for a correct factual answer and two points for a correct explanation.

Mean scores for each group on each subtest are shown in Table 1. An analysis of variance of these scores revealed no significant main effect of either question-asking or hints. Nor was there any significant interaction between subtest and either of these variables, despite the fact that subtests II and III were specifically designed to simulate the question-asking and hints given during training to certain groups. The only significant effect ($P<.01$) was the second-order interaction, subtest x question-asking x hints. The means shown in Table 1 do not offer any clear, sensible interpretation of this interaction. Note that the criterion test differences correspond somewhat to differences on the test over the geography unit taken just prior to the experiment (see Table 1). We had intended to use test scores on this previous unit as a control variable in an analysis of covariance, but the regressions were too heterogeneous.

It can also be seen from Table 1 that the groups (KQ and KW) which reviewed principles instead of practicing problem-solving did just as well on the criterion application problems as did the groups who had practiced such problem-solving. Just as remarkably, the KQ and KW groups did no better on the Part IV of the test which required only recognition of the very facts and principles reviewed by these groups. In sum, there was no evidence that practice at either (a) question-asking, (b) the use of hints, or (c) review of principles, was any more beneficial for a criterion subtest emphasizing such a problem-solving subskill than for criterion performance in general.

Nor was there any general advantage to asking questions or to getting hints during training (no main effects), given that the time required to do these things could be spent on other relevant learning. The equivalence of study time becomes especially important when we note that question-asking groups worked on less than half as many problems during training as the work-alone groups (mean number of training problems attempted: PQH 8.21, PQ 7.38, PWH 18.50, PW 18.21). Since all groups were given the same time period for training, it would seem that lack of training at any particular subskill (of those examined here) was compensated for by emphasis on some other aspect of problem-solving.

The generally low scores on the criterion test (maximum possible score = 80, overall mean = 39+) are fairly typical for posttests emphasizing application problems. Still there is plenty of room to believe that more

Table 1

MEAN TEST SCORES

	CRITERION: Part I		CRITERION: Part II		CRITERION: Part III	
	Question Groups	Work-alone Groups	Question Groups	Work-alone Groups	Question Groups	Work-alone Groups
Problems and Hints	9.29	10.13	12.43	10.58	6.79	7.04
Problems, No Hints	10.50	10.12	9.04	13.29	6.96	10.46
Principles Review	12.13	12.67	10.69	15.42	8.25	8.00

	CRITERION: Part IV		TOTAL CRITERION SCORE		PREVIOUS GEOGRAPHY UNIT	
	Question Groups	Work-alone Groups	Question Groups	Work-alone Groups	Question Groups	Work-alone Groups
Problems and Hints	10.36	9.25	38.87	37.00	18.33	18.00
Problems, No Hints	9.69	10.86	36.19	44.73	17.27	18.62
Principles Review	10.13	10.33	41.20	46.42	16.38	22.80

effective training for problem-solving could be devised. The low scores may reflect in part the observed hastiness of the Ss. Although urged to work carefully, most Ss breezed through the tests pretty rapidly and had finished each part in about half the allotted time.

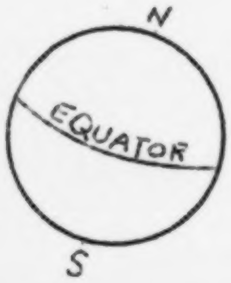
Among the problem training groups, both question-asking ($P<.10$) and hints ($P<.05$) during training appeared to facilitate question-asking slightly during Part III of the criterion test. Group PQH which had both question-asking and hints during training asked more than twice as many questions (per S) on the test as the other groups combined. So few questions were asked even by this group, however (1.8 questions per S) that one would expect this to have little direct effect on test performance.

The teachers volunteered the informal observation that more questions were asked by the class as a whole after the experiment than before. Although there was no evidence that better problem solving was an immediate correlate of greater question-asking during training, it is conceivable that changes in strategy resulting from the use of hints and question-asking would only reveal their effects over a longer time period than that of the present experiment.

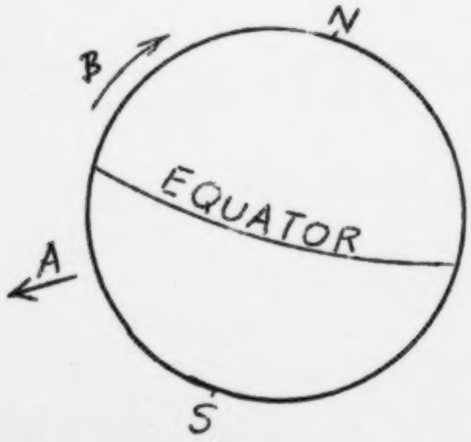
APPENDIX A

Basic Program on Horse Latitudes (2 sample pages)

1. Look straight up in the air. You are looking toward the (center of the earth, sky, north, south).	1. sky
2. When you look down, you are looking toward the (center of the earth, sky, north, south).	2. Center of the earth
3. The N and S on the map below stand for the - ? - and - ? - poles.	3. north, south



Use the diagram below in answering Nos. 4 and 5.



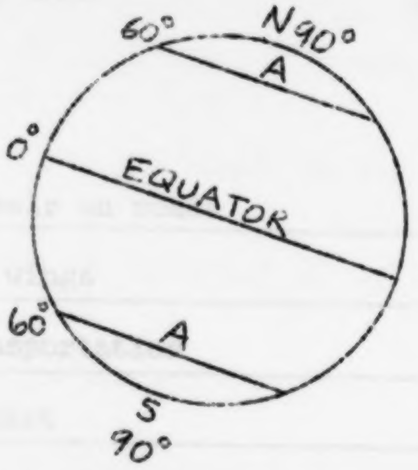
4. Arrow A points (north, south, down, up).	4. up
5. Arrow B points (north, south, down, up).	5. north

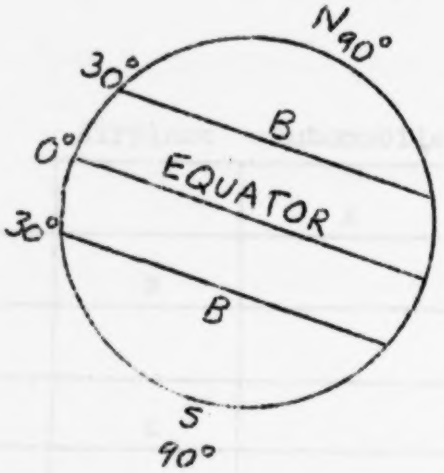
46. The air falling back to earth one-third of the way to the poles is (moist, dry).

46. dry

47. The air that rises at the equator falls back to earth at (A, B).

47. B


60°
N 90°
A
0°
EQUATOR
60°
A
S
90°

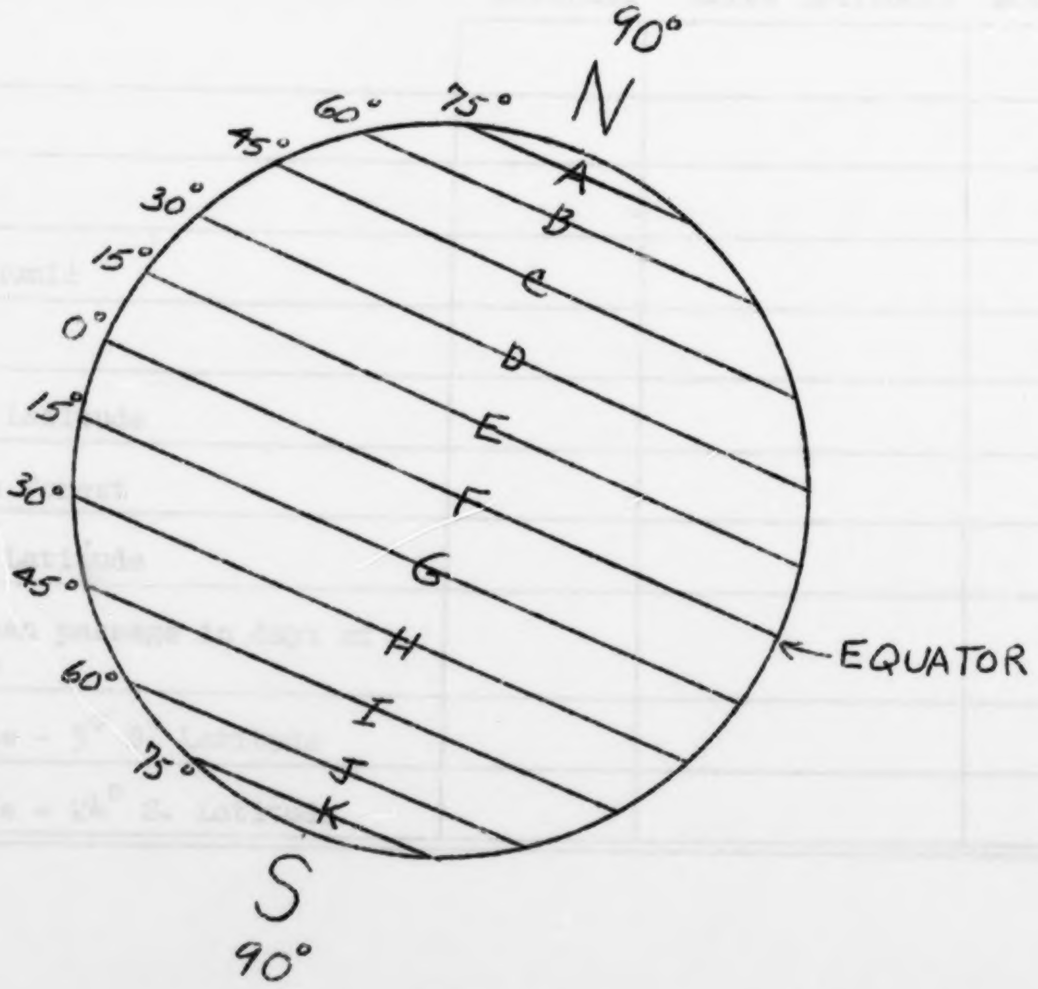

N 90°
30°
B
0°
EQUATOR
30°
B
S
90°

48. How many degrees from the equator are the horse latitudes?

48. 30°

49. Which are the horse latitudes? (A, B, C, D, E, F, G, H, I, J, all, none)

49. D, H


90°
N
60°
75°
45°
30°
15°
0°
15°
30°
45°
60°
75°
A
B
C
D
E
F
G
H
I
J
K
EQUATOR
S
90°

APPENDIX B

Review of facts and principles given to groups KQ and KW

MATCHING GAME

Sample:

	Airplane	Automobile	Both	Neither
a travels on roads		x		
b. has wings	x			
c. transportation			x	
d. fastest	x			
e. has a smoke stack like a train				x

	Doldrums	Horse Latitudes	Both	Neither
1. Air rising				
2. Air falling				
3. Much wind				
4. air hot and humid				
5. desert				
6. 25° - 35° N. Latitude				
7. tropical rain forest				
8. 25° - 35° S. Latitude				
9. Difficult ocean passage in days of sailing ships				
10. 5° N. Latitude - 5° S. Latitude				
11. 6° S. Latitude - 24° S. Latitude				

Tell Why:

1. It is incorrect to say, "I went up to Alaska." (A p. 33)
2. It is incorrect to say, "I went down to Mexico." (A p. 33)
3. It is incorrect to say, "Canada is above the United States." (A p. 33)
4. It is incorrect to say, "Australia is below Asia." (A p. 33)
5. The air high over the doldrums is cold. (B p. 33)
6. The air that rises at the doldrums falls back to earth. (I p. 5)
7. In the horse latitudes the air is dry. (B p. 33)
8. In the horse latitudes there is little or no wind. (B p. 33)
9. The 30th parallels are 1/3 of the way to the poles. (B p. 46)
10. The horse latitudes were a difficult ocean passage in the days of sailing ships. (B p. 33)
11. Most of the world's deserts are near the 30th parallels. (C pp. 33-34)

APPENDIX C

Instructions and first two pages of the Problem 1 booklet for each experimental group given practice problems.

Group PWH

INSTRUCTIONS

1. Take your time and think about each problem carefully. You don't have to work all the problems. Stay with one problem until you have learned all you can about it, then write out a complete answer on the back of the booklet. You only get one try at answering each problem. After you have written your answer, raise your hand. A teacher will check it for you.

2. Use your workbook, summaries, atlas, globe, encyclopedia, etc. to help you solve the problems.

3. If you have reviewed your workbook and summaries carefully and still can't solve a problem, tear the problem page loose from the staple in the middle. The hint on the next page may give you some ideas. USE AS FEW HINTS AS YOU CAN, but be sure you can solve the problem before writing your answer.

Group PWH

Name ______________________

Problem 1

Whis is heavier out in the open, a cubic foot of cold air or a cubic foot of warm air? Why? (Write answer on back of booklet).

DO NOT GUESS. Review your program summaries, or atlas until you are sure you can answer the problem correctly. If you get stuck and need a hint, tear this page loose from the staple above and see next page.

---

page 2

Hint 1: When air is heated the molecules bump into each other and bounce farther apart.

Think about the problem again before you look at the next hint page.

Group PQH

INSTRUCTIONS

1. Take your time and think about each problem carefully. You don't have to work all the problems. Stay with one problem until you have learned all you can about it, then write out a complete answer on the back of the booklet. You only get one try at answering each problem. After you have written your answer, raise your hand. A teacher will check it for you.

2. Learning to ask questions about problems may help you learn to solve them. Ask the teachers as many questions as you can think of about a problem before writing your answer to the problem. You must ask at least 3 questions, even if you already know the answer to the problem. Whenever you think of a question raise your hand, and a teacher will come and answer it for you. If all teachers are busy, write the question down and think of another question while waiting. You may use your workbooks, maps, and anything else you want.

3. If you have asked at least one question and can't think of another, tear the page loose from the staple in the middle. The hint question on the next page may give you some ideas. Teachers will answer the the hint questions as well as your own questions.

Group PQH

Name ______________________________

Problem 1

Which is heavier out in the open, a cubic foot of cold air or a cubic foot of warm air? Why? (Write answer on back of booklet).

Your questions: (ASK 3 OR MORE QUESTIONS OF YOUR OWN BEFORE ANSWERING THE PROBLEM)

Ask at least one question before tearing this page loose from the staple above to see the hint question on the next page.

- -

Page 2

Hint question 1: What happens to air molecules when they are heated?

Your questions: (Ask at least one question of your own before looking at hint question on next page.)

Group PQ

INSTRUCTIONS

1. Take your time and think about each problem carefully. You don't have to work all the problems. Stay with one problem until you have learned all you can about it, then write out a complete answer on the back of the booklet. You only get one try at answering each problem. After you have written your answer, raise your hand. A teacher will check it for you.

2. Learning to ask questions about problems may help you learn to solve them. Ask the teachers as many questions as you can think of about the problem before writing your answer to the problem. You must ask at least 3 questions, even if you already know the answer to the problem. Whenever you think of a question raise your hand, and a teacher will come and answer it for you. If all teachers are busy, write the question down and think of another question while waiting. You may use your workbooks, maps, and anything else you want.

3. The extra sheets in each problem booklet are scratch paper for you to make notes on and write questions on.

Group PQ

Name ______________________

Problem 1

Which is heavier out in the open, a cubic foot of cold air or a cubic foot of warm air? Why? (Write answer on back of booklet).

Your questions: (ASK 3 OR MORE QUESTIONS OF YOUR OWN BEFORE ANSWERING THE PROBLEM)

- -

Page 2

(blank scratch paper)

Group PW

INSTRUCTIONS

1. Take your time and think about each problem carefully. You don't have to work all the problems. Stay with one problem until you have learned all you can about it, then write out a complete answer on the back of the booklet. You only get one try at answering each problem. After you have written your answer, raise your hand. A teacher will check it for you.

2. Use your workbook, summaries, atlas, globe, encyclopedia, etc. to help you solve the problems.

3. The extra sheets in each problem booklet are scratch paper for you to make notes on.

Group PW

Name ______________________________

Problem 1

Which is heavier out in the open, a cubic foot of cold air or a cubic foot of warm air? Why? (Write answer on back of booklet).

---

page 2

(blank scratch paper)

APPENDIX D

First ten practice problems given to all problem training groups and hints given to groups PQH and PWH

1. Which is heavier out in the open, a cubic foot of cold air or a cubic foot of warm air? Why?

(Hints) a) When air is heated the molecules bump into each other and bounce farther apart.
b) The more molecules there are in a cubic foot of air, the more that cubic foot of air weighs.
c) Cold air molecules are packed closer together than warm air molecules.
d) There are more air molecules in a cubic foot of cold air than in a cubic foot of warm air.

2. You would go in what direction from the equator to the South Pole?

(Hints) a) North means toward the North Pole.
b) South means toward the South Pole.

3. There is little rain at the horse latitudes because there is (so much, so little, no) rain in the doldrums. Why?

(Hints) a) Air at the doldrums rises very high.
b) When air cools it loses its moisture.
c) As air rises to higher altitude it cools more.
d) There is very little water vapor at high altitudes.
e) Air flows from the doldrums to the horse latitudes at very high altitude.
f) Most of the high altitude air from the doldrums falls to the earth at the horse latitudes.

4. About how many degrees are the horse latitudes from the poles?

(Hints) a) The North and South Poles are at 90° latitude.
b) The center of each horse latitude belt is at about 30° latitude.
c) The horse latitudes are 30° from the equator, which is 0° latitude.
d) To go from 30° latitude to 90° latitude you have to cross 60° of latitude.

5. In the Kalahari Desert (Latitude 27° S.) in South Africa, is the air rising or falling? Is it warm or cold? Is it moist or dry?

(Hints) a) The horse latitudes extend from about 25° to 35° latitude (N. and S.)
b) Most of the dry air flowing north and south from the doldrums at high altitudes falls back to the earth at the horse latitudes.
c) As air falls it warms because of compression.

6. Where would it be easier to sell clothes-drying machines, in the doldrums or the horse latitudes? Why?

(Hints) a) The machines would probably sell best in lands where it takes clothes a long time to dry outdoors.
b) Things become dry as water evaporates from them.
c) Water evaporates faster in warm dry air than in cool or moist air.
d) Air in the doldrums is moist.
e) Air in the horse latitudes is dry.

7. Air moving from the floor to the ceiling of a room is likely to be (warm, cold) air. Why?

(Hints) a) Cold air and warm air do not weigh the same.
b) Cold air is heavier than warm air.
c) A heavy gas sinks to the bottom of a container while a lighter gas rises to the top.

8. The horse latitudes and doldrums shift north and south as the seasons change. Central Mexico is in the horse latitudes in winter but is nearer the doldrums in summer. Does it rain there more in winter or summer? Why?

(Hints) a) Air falling at the horse latitudes remains dry.
b) Air rising in the doldrums is moist.
c) When air rises very high it loses its moisture.
d) The doldrums get much more rain than the horse latitudes.

9. In a cold country, it is better to plant apple orchards on a hillside than on flat ground, because they are not as likely to freeze when they are in blossom. Why is this?

(Hints) a) Cold air is heavier than warm air.
b) Heavy air sinks down while light air rises.

10. Air in the doldrums and air in the horse latitudes move
A. at right angles to each other.
B. in the same direction.
C. in opposite directions.

(Hints) a) Air in the doldrums is rising.
b) Air in the horse latitudes is falling.
c) Up and down are opposite directions.

Chapter III

Laboratory Study of Individual Performances on Programed Learning Tasks

Lloyd O. Brooks

Optimal Difficulty

This chapter is a report of laboratory study of the performances of individual students. Priority was given to methods of analytical study rather than the devising of techniques which incorporate what is currently known about thinking into what is currently known about programed instruction. Regardless of what new techniques of programing are developed which favor critical and thoughtful learning, transfer or generalization to new situations, and persistence on challenging conceptual tasks, it will be necessary to detect the effects of these techniques. By studying, in detail, existing programs, through student performances, those portions may be discovered which are best with respect to certain desirable characteristics--such as level of difficulty. New techniques of programing may evolve out of exaggerations of these characteristics which seem to be related to beneficial effects.

We may look for such effects in gross or subtle student performances. While gross performances, such as scores on an end-of-program criterion test, may be of considerable practical and theoretical significance, the details of performance throughout the program are likely to guide better efforts at revision and program improvement--or even what is likely to be best for the student to do next in progressing through the program. This is especially true as we learn what details to look for and how best to interpret our observations. Knowledge of both kinds of performances are complementary.

Skinner (1958) wrote, "Whether good programing is to remain an art or to become a scientific technology, it is reassuring to know that there is a final authority--the student. An unexpected advantage of machine instruction has proved to be the feedback to the programmer" (p. 18). During the past six years remarkably little work has been addressed to identifying precisely the kinds of feedback from this final authority which best offer guidance for the programer.

It is clear from Skinner's report that errors were an important kind of feedback. Skinner also noted that, "The time required to complete a disk [a 29-item or so unit of programed material] depended upon the difficulty. There are also wide individual differences among students" (p. 42). Continuing to express concern for the role of errors during learning, Skinner (1961) refers to the work of Herbert S. Terrace in the Harvard laboratories--work later reported and apparently continued by Terrace while at Columbia University--as "a notable example of the power of effective programing" (pp. 93-94). This work serves to elaborate on the importance of errorless performance by showing that what is learned without errors during learning is superior to learning involving the correction of errors in terms of both dependability of learned performance under changed conditions and, perhaps relatedly, freedom from unpleasant emotion such as anxiety (Terrace, 1963 a, b, and c).

The data of one of his studies suggested to Terrace "that once errors occur during the formation of a discrimination, subsequent performance is permanently affected on that, and also on related discriminations. This generalization should prove to be of prime importance in analyzing discrimination performance. It is also directly relevant to programing techniques for teaching machines in that it provides evidence for the importance of learning a program with the fewest possible errors (c.f. Skinner, 1958) (Terrace, 1963, b, p. 231).

(In anticipation of the possibility that errors are, indeed, to be avoided, and in view of the benefits of diagnostic testing which seems to require errors, a means of errorless testing will be considered later in this report.)

Perhaps the issue of optimal error rate as an issue pertaining to this portion of the report can be explained as follows. Several theoretical arguments questioning the value of errors could be cited, notably in the writings of Skinner. The possible advantages of errors under some conditions could be examined in some detail, following such leads as suggested by Amsel (in Lumsdaine and Glaser, 1960). Undoubtedly other observations similar to this by Ferster and Sapon could be assembled: ". . . we discovered that level of difficulty was the biggest determinant of the student's sustained motivation. Whenever the level of difficulty increased too rapidly from unit to unit, so that large numbers of errors were made, the reaction was uniform in nearly all the students, who reported being emotionally upset and who reported a strong tendency to stop even though they may not have completed all of the work" (in Lumsdaine and Glaser, 1960, p. 183). Such statements led to the position in the present study that there seem to be several reasons for avoiding errors during most learning. Moreover, some progress seems to have been made in giving new meaning to a suggestion by Skinner that perhaps the advantages afforded by errors can be realized in other ways at less hazard.

The development of present-day programed instruction seems to have been distorted by a methodology which is effective mainly in keeping a program from becoming too difficult. Few current practices keep it from becoming too easy. This one-sided constraint may be justified on the grounds that a good program is easy all the way but does deal with difficult material. The student is merely spared the effects of difficult material by being gradually prepared so that his competence dispels the difficulty of advanced material.

It is also possible that too much difficulty is being removed from the student's work along the way. Perhaps material should not be so difficult that the motivated and properly-prepared student makes errors, but perhaps this student should not be kept at the safe distance from error which results when a conscientious programer becomes too exclusively preoccupied with only error data as his feedback from the final authority--the student. Too little difficulty can also have an adverse effect on motivation as the unchallenged student becomes bored and inattentive.

A possible failing suspected of programed learning is that while the subject matter of the program may be taught better than ever before, the student develops less skill in applying what is learned when circumstances are changed. It seems plausible that generalization or transfer is improved by some of the factors which interfere with maximal efficiency as far as subject matter is concerned. Imperfections in the learning circumstance may favor the development on the part of the student of special skill in thinking about new material in a variety of ways, organizing, and reviewing effectively, or supplying to himself supplementary examples of a familiar nature.

The response to this possibility would seem to be to learn more precisely what kinds of challenges to learning are of greatest benefit and more detail concerning how this benefit comes about. It would seem that whatever it is that is of benefit in developing high-order mental skills could be arranged for in a programed way just as academic or technical subject matter is programed.

A very reasonable approach is a direct approach. Whatever seems to be beneficial to transfer, generalization, problem solving, or critical thinking can be put in program form and the program revised with the aid of error data, student comments, and criterion test performance.

An alternative employed in this phase of the study has been to study some new kinds of measures which supplement such data as error rate in an effort to be in a stronger position for later developing programs with new, and better defined, characteristics. The question also arises as to whether programing really does clash with the development of transfer, or generalization, problem solving, or other desired outcomes. Also, what portions of a program are best or worst in such respects.

As concern increases with whether programed learning fails to do some things as well as learning which is less carefully arranged for, it becomes increasingly important to examine in greater detail ways in which students perform in relation to the programed situation.

In defense of the strategy used in this phase of the work, it can be pointed out that, in compensation for deferred work on what may be regarded as some of the central issues--such as problem solving, generalization, or originality--the findings and implications of this phase of the project, being basic and analytical in nature, bear on programing issues in general--regardless of program content.

Response Latency

For some time there has been dissatisfaction with error rate as the major descriptive index of program excellence. Error rate can be low because of overprompting. A criterion test tells much more about the effectiveness of a program but less about which specific portions are too easy or too difficult. Granted that prompting is not excessive and that prompts are eventually vanished, error data tell more about items which are too difficult than they do about items which are too easy.

On the other hand, latency implies something about the relative difficulty of items which are not missed. Often items could be further complicated or made to teach more if the programer only knew how excessively easy they were. Another advantage of information concerning latency is its greater efficiency, in contrast to error data, in telling us something about every item.

The kind of latency referred to here is the time from the presentation of a question until the student answers the question. Some readers may be more inclined to regard latency as the time between a smaller unit of the stimulus, such as a word or letter, and a response. There may be a preference for the term "reading-and-answering time" rather than "latency." A further complication is that whereas the interval, which is referred to as latency, clearly begins with the presentation of a frame of programed material, it ends with the response the student makes in exposing the program answer, a response which usually occurs a fraction of a second after his answer. Nevertheless, the approximation is sufficient to regard this particular time interval as essentially the latency of a discriminative operant or instrumental response and thereby to gain the conceptual benefits of a term used by several major learning theorists as one index of response strength.

A previous unpublished study (Brooks, 1961) has confirmed the validity of this interpretation in the context of programed learning by showing, under a variety of conditions, that there is a relationship between error rate on composed-answer items and latency. Students typically spend more time answering questions that are missed, questions that are more difficult. Furthermore, the correlation between latency defined both by time from question-to-answer or question-to-confirmation justifies any technical irregularity in the use of this term.

Suppes (1964) writes, "Only during the last year or two has there been a serious revival of interest in response latency as a variable that may serve as a central criterion of learning. Some of this recent experimentation is showing clearly what all of us probably know in a muddle-headed way: namely, that, for many kinds of learning, latencies are a much more sensitive index than are response errors themselves. For at least three areas of the elementary-school curriculum--reading, mathematics, and foreign language--there is no doubt that response latencies are more sensitive measures of skill mastery and depth of learning than the responses themselves" (p. 87).

Siegel (1964) calls attention to response time considerations in the case of concept formation and gives the results of an exploratory study. Siegel explains partly the lack of latency studies with this comment, "Engineering or instrumenting the objective measurement of decision, judgment, discrimination, or response times is relatively expensive" (p. 435). A contribution of his paper is the description of ". . . a not too expensive instrument of relative simplicity and dependability that will permit the objective recording of response latencies within a discrete trials procedure" (p. 435). In addition to concept-formation tasks, Siegel reports, "This measure is being investigated currently in this laboratory in the context of reversal-nonreversal transfer (Kendler and Kendler, 1962) and in other than two-category tasks" (p. 441).

Suppes cites the work of Lambert (1955) as an example of one of the few isolated studies of response latencies in second-language learning. This work deals with the measurement of "fluency" in bilinguals. Suppes comments, "However, in view of the obvious relevance of response latencies to measures of second-language mastery, the paucity of serious experimental studies in the literature is almost shocking" (p. 33).

Mager has noted the potential usefulness of latency data in assessing varient forms of restatement of program items when a well-developed program in one language is then translated into another.[1]

Suppes cites an example of one of the ways in which response latencies have proven useful as sensitive measures of learning. Russian words were spoken by a native speaker with Cyrillic response alternatives in multiple-choice format. After hearing a word, S pressed a key corresponding to one alternative and was immediately informed of the correct answer. Suppes reports, "We found that latencies strongly depended upon the type of item. For example, longer latencies were associated with items having Cyrillic characters similar to each other and distinct from Roman letters. It was possible to divide the items into five categories and have for all sessions a clear separation of the latencies. Most important, even after the response-latency curves had flattened out to an asymptotic level by essentially the tenth session, the clear separation in response latencies for the five types of items remained. A foremost problem for language teaching is to devise methods for breaking through these asymptotic levels of response latency by finding methods that will, with sufficient practice, drive them down to the asymptotic latency levels of the subject's native language" (p. 88).

In the present study as well as in the previous one (Brooks, 1961), cited previously, lessons were repeated and the correlations between first and second presentations of lessons considered. In the context of programed learning, latency data tell far more than error data when material is repeated. Many interesting questions are raised such as why students spend longer on some items the second time they are presented than the first, and whether the degree of drop in latency of other items suggests, in a reliable and valid way, the degree of their mastery.

A useful finding in early study of latency data in the context of programed instruction pertains to program revision and points up one basis for criticism of the too-exclusive preoccupation with error data. Sometimes frames are revised and made easier when the fault lies in earlier treatment of the topic covered by these frames. Each time these, or incorrectly-identified earlier frames which are assumed to be the cause of the difficulty, are revised, unnecessary simplification, and often length, accrues to the program. Long latencies on earlier portions of related material may signal the points most in need of revision, especially with benefit if these points deal with essential prerequisites.

---

[1] Mager, R. F. Personal communication.

Another suggestion in the earlier study (Brooks, 1961) was that latency data were more predictive of error rate for the more difficult items, such as frequently-missed items on a review test. In part, this seemed to be true because the time spent on easier material was more largely a function of amount of material. Of importance to the present study is the imperfection, in the sense of non-perfect correlation, of the relationship between error rate and latency--even when the refinement is introduced of treating latency as a kind of residual which remains after allowance is made for amount of material.

The imperfection in the relationship suggests that there are possibly two kinds of difficulty. One is best indexed by latency, the other by error data. This notion is of considerable significance to the present work and germane to the topic of prompting, especially to the discussion and findings of Angell and Lumsdaine (1961). By prompting in a manner that increases difficulty, as defined by latency, without increasing, or by increasing at a much lower rate, error tendency, it should be possible to cue at some near-optimal point.

The nature of some of the possible motivational benefits is suggested in the earlier writing of Skinner (1953) where he speaks of one reinforcing consequence in making a decision as simply escape from the aversiveness of indecision (p. 244).

Increasing the difficulty of items to the point of greatest challenge should afford thestudent who answers correctly greater satisfaction with his achievement as well as terminate his awkward moment of being unsure.

Other benefits which are not totally of a motivational nature are suggested by the work of Gagne and his associates (1962 and more recently), Kersh (1963), and by Wittrock (1963). These have to do with degree of guidance or so-called discovery learning. It seems reasonable that maximal suberror difficulty would favor the most efficient shifting from responding under the control of prompts to the critical properties of stimuli. Gagne has expressed interest in the possible relevance of latency data to such matters as optimal guidance, suggesting at the same time that greater benefits might be expected for some subject matters than others (e.g., mathematics).[1]

In both the early study (Brooks, 1961), cited previously, and the present study, the time spent on a single frame of programed material was further divided so that a separate measure was obtained of the time from machine presentation of the program answer to the student until he advanced to the next frame. It was typical for the student to spend more time examining a program answer when it did not agree with his own answer. It was considered possible that a change in this respect would occur after many sessions were spent working on a program and that this change would be correlated with other evidence of decreased student interest. Since student interest, as casually observed, did not change--decrease--in a way which favored evaluating this possibility, it will have to be reconsidered when a means is devised for manipulating student interest. However, some relevant observations are possibly potentially available in some of the data of the present study.

---

[1] Gagne, R. M. Personal communication.

It should be recognized that a major objective of this investigation has been to demonstrate the general usefulness of a method of experimental analysis. In this attempt there has been, perhaps, disproportionate concern with R-R (response-response) laws or relationships, one dependent variable in relation to another--latency to error. One important independent variable--amount of material--was studied statistically rather than manipulatively. Rarely were special instructions or restrictive conditions employed.

The intention was to develop an "apperceptive mass" on the part of the experimenter with respect to his various naturalistic observations under which objective measures were obtained. As points of interest developed, restrictive conditions or instructions were introduced to serve a particular purpose. As study progressed, manipulation of certain variables, such as delay of opportunity for scoring, did occur.

Procedures

Subjects

Three 5th grade students, one high school junior, 6 high school seniors, 2 college freshmen, and one college junior participated in the study. With the exception of the two college freshmen, the subjects were females.

Previous stratified and quite diverse random sampling had not suggested reasons for defining a particular population for the kind of study contemplated. When the writer conducted the 1961 study, students were selected for study in terms of those first to finish earlier material, those about average, and the slowest. Latency-error relationships seemed to hold for all Ss. Less systematic study since then has not pointed up any particular sampling considerations. However, since the present study was concerned with the topic of difficulty, many of the student subjects included in the present study would be inappropriate users if one were interested, say, in using the error data to revise the program or to assess its probable effects with a college population.

Ss were volunteers. They were given an honorarium on the basis of hours of service, with the exception of one brief special-incentive session.

Materials

A program dealing with earth-sun relationships being developed by P. L. Campeau was the only program used by the youngest subjects. Several of the high school students provided data useful to the present study and to revision of a first draft.

Nine of the subjects provided some data from a program dealing with verbal problems in algebra.

The Analysis of Behavior by J. G. Holland and B. F. Skinner (1961) provided a better format and the possibility of learning something about the role of exhibits, or panel materials, as well as being at least marginally suitable for both advanced high school seniors and college students. It also provided review tests. This program, already the object of some study, seemed sufficiently classical in the development of programing to warrant further study. Accordingly, with the permission of the publisher, McGraw-Hill, a filmed version of the entire program was prepared. Of the nine subjects using this program, one worked only the first set, another, 8 sets, with the other subjects completing approximately the first one-fourth of the program, 16 to 20 or so sets. Most of these subjects worked through this material twice.

Six of the older students worked through the first 30 or 60 frames of "Reflection and Rotation--Symmetry of Two and Three Dimensional Objects" by A. L. Loeb and E. Haughton.

Thirty-five isolated, factual statements in a program-like form, prompted and composed to create three levels of difficulty by R. A. Goldbeck (Goldbeck and Campbell, 1962), were used late in this study with 7 subjects.

A set of 30 items dealing with latitude was prepared to use in studying different techniques of revision. It was developed late in the study by the writer, and the first draft was worked through twice under special incentive conditions by 7 subjects.

Apparatus

Two DuKane viewers using 35 mm. filmstrips and providing for composed answers, and one Mast teaching machine were connected to latency-recording equipment.

During this study only one of these devices was in use at any one time. Each student worked alone in a small room where he and his work could be observed by E from an adjacent room through a one-way view window.

A Harvard C-3 cumulative recorder received a double-pulse per second as response input, thus constructing one millimeter of ordinate value for each second S spent reading and answering a program item. Each button press reset the stepper of this recorder, and reset was automatic for reading-and-answering time followed by another ordinate representing the time spent observing the program answer and in advancing to new material. The advantage of this recording technique was that it proved trouble-free and provided the safeguard that there would be noticeable irregularities in the record if the pulse generator (a synchronous motor providing cam action on a microswitch) were to fail.

Later, and incidentally, the dependability of pulsing assured, a three-counter printing counter was introduced to provide latency data in numbers rather than in lines which had to be measured. At this point, the cumulative recorder was used to provide a cumulative record of frames completed.

Instructions

S was told that characteristics of the materials were of interest to E, that we could learn only about some characteristics through S's performance, and that it was most important that he answer each question correctly. However, the time spent on each question was also of interest; and, therefore, S should work from the beginning to the end of each lesson as quickly as possible while spending whatever time it required on each question to answer correctly. S was encouraged to guess about answers if necessary, to avoid leaving any blank, but to draw a line if even a guess were not possible.

Special instructions intended to favor motivation were given late in the study in the case of the set of items dealing with latitude. S was told that whereas always before we had been interested in characteristics of the material, we were now trying to find out which students could do best in dealing with difficult material. If none of the questions were missed S would earn a 50-cent bonus.

In a still later phase when special modes of machine presentation were employed to study delay in feedback, E mentioned that there was interest in which student was most capable of dealing with difficult material when the task was made even more difficult by the peculiar functioning of the machine. Under these conditions S responded to the isolated-facts materials, and encountered delay of answer for about 10 seconds after the response on half of the items. An incidental finding, due to improper circuit design early in this part of the study, suggested some interesting possibility for further study.

Results

Latency Related to Error

The results presented in Table 1 and Table 2 are for a first and second presentation respectively. For each set of frames in The Analysis of Behavior completed by each S the median read-and-answer time for frames on which there was at least one error (some frames had several blanks--possibilities for error) is presented and compared with such a median of frames on which there was no error.

These results add to the generality of the previous findings based on remedial English programed material (Brooks, 1961). For each S the time spent on frames that were missed was greater in most instances.

This result is based on all of the data. In the data of Tables 1 and 2 no attempt was made to exclude frames which were associated with longer times as a result of S reading related panel material. Sets with the asterisk had panel exhibits.

If we consider the first-presentation data for each of the 8 Ss in terms of the number of instances in which median times based on incorrectly answered frames are greater, it is evident that there are more such instances for each S and that $P<0.01$ for a sign test.

The number of frames on which there was at least one error is given as the numerator of the fractions in Table 3. The denominator indicates the number of frames each S completed; this defines the opportunities for error. Table 3 can be used in interpreting the number of cases on which the medians of Table 1 are based. Table 3, rather dramatically, points up how inappropriate this material was for certain Ss. Those Ss whose data are given in the left columns lack the background assumed for this program. From left to right there is a change from highly inappropriate to highly appropriate program users.

Table 4 gives results similar to Table 3 but for a second, rather than first, presentation of a lesson. In most tables, Set 5 is used as if it were two separate sets in an effort to consider changes from early to late performance on relatively homogeneous material. S #3 seemed to make two more errors on the first than second half of Set 5 on the first presentation but one more on the second half on the second presentation. Note how S #3 seemed to perform, in terms of errors, on a second presentation occasionally about as well as the best S on the first presentation, e.g., consider Set 8.

Table 1

First Presentation Median Read-and-Answer Time for Frames of

The Analysis of Behavior With and Without Error

Median (in Seconds)

Set	S #1		S #2		S #3		S #4		S #5		S #6		S #7		S #8	
	Error	No Error	Error	No Error	Error	No Error	Error	No Error	Error	No Error	Error	No Error	Error	No Error	Error	No Error
1	34.5	20.5	25	19.5	22.5	12	24	16	23	14	30	22	20.5	14	30	19
2	24	22.5	22	23	24	26.5	16.5	22	21.5	21	25	22	29	15	27.5	21.5
* 3	38	37.5	31	29							29.5	13	20	16		
* 4	25	28	33	20.5	33	14.5	23	16	35	15.5	30	18	8	13	39.5	22
(1st 1/2) 5			27	23	37	14.5	15	15	29.5	21	19	22	18	15	20.5	20
(2nd 1/2) 5	29	26	24.5	22	35.5	22	33	23	36	26.5	27	22	14	17	22	25
* 6	24.5	25	39	18.5	25	25.5	31	22	32	14.5	42	17	42	14.5	71	22
7	21	17	19	15	24.5	15	30	11	20	16	21.5	15	21.5	12.5	13	13
* 8	35	28	32	32	35	27	29.5	20	39	24.5	41.5	26	18	18	23.5	20
9	27	26.5	23	19.5	35	22.5	30	24	40	24	28.5	24	21	19		
10	22.5	18.5	25.5	19	32	23	19	21	21.5	26	24	24	15.5	13.5		
* 11	25	25	32	29	34	28	32	18	33	28	34	25	18	15		
* 12	29	29	43	30	23.5	23	31	19	35.5	22.5	17	26	22	18		
* 13	29	25	28	22	24.5	25	18.5	20	25.5	23	37.5	21	15	14		
* 14	28.5	23	28	21.5	45	24	24	18	27.5	23	32	22	25	15		
* 15	27.5	37	34	31	42.5	23.5	17	20	36	30	23.5	24	21	16		
16					28	32	23.5	26	43	25.5	31.5	24	21	14		
17					41	27	22.5	28	38.5	30	29.5	30	22	17		
* 18					40	29	24.5	20.5	30.5	22	32	25	28	16		

* Set had panel exhibit

Table 2

Second Presentation Median Read-and-Answer Time for Frames of

The Analysis of Behavior With and Without Error

Median (in Seconds)

Set	S #3		S #4		S #5		S #6		S #7	
	Error	No Error	Error	No Error	Error	No Error	Error	No Error	Error	No Error
1			17	12			23	13	16	10
2							16	9	11.5	9
* 3	51	31	19.5	26	28	21	17	15	8	11.5
* 4	50.5	23	30.5	17	39	21	19	17	28	13
(1st 1/2) 5	28.5	20	17	17	19	17.5	25	18	12	11
(2nd 1/2) 5	32	20	26	13.5	35	21	25	21	8	13
* 6	27	19	20.5	15	34.5	17	22	15	17	11.5
7	58	15	17	11	28	14	16.5	12	14	9
* 8	32	26	25.5	17	32	21.5	56	18.5	15	12
9	22	22	16	14.5	24	18	20	23	12	13
10	28	16	15	13.5	14.5	15	29	19.5	13	11
* 11	31	24	21	16	32	19	23	17.5	13.5	12
* 12	66.5	24.5	19.5	16.5	42	20	15	18	17.5	16
* 13	40	22	15	14.5	24	17	20	17	14	12
* 14	31	21	23	16.5	51	19	27	17	12	11.5
* 15	54.5	24	26.5	19.5	35	19	19	21	14.5	13.0
16	31	20.5	21	24	—— **	22	26	21	9	11.5
17			33	22.5	22	21	18	23	17.5	14

* Set had panel exhibit ** S answered all frames correctly

Table 3

The Number of Frames with One or More Errors Divided by the Number of Completed Frames

(First Presentation)

Set	S #1	S #2	S #3	S #4	S #5	S #6	S #7	S #8
1	32/54	38/54	8/29	28/54	5/29	7/54	11/54	3/54
2	20/30	21/30	20/30	18/30	8/30	5/30	5/30	2/30
* 3	21/29	23/29	——	——	——	5/29	5/29	——
* 4	25/29	25/29	11/29	17/29	11/29	4/29	3/29	4/29
(1st 1/2) 5	23/29	18/29	17/29	12/29	4/29	5/29	3/29	4/29
(2nd 1/2) 5	17/30	22/30	15/30	15/30	8/30	7/30	6/30	5/30
* 6	22/29	15/29	11/29	7/29	7/29	4/29	1/29	2/29
7	15/30	15/30	12/30	7/30	11/30	5/30	2/30	1/30
* 8	24/31	17/31	16/31	8/31	9/31	8/31	4/31	8/31
9	23/29	23/29	9/29	10/29	3/29	14/29	10/29	——
10	24/32	22/32	9/32	7/32	8/32	8/32	6/32	——
* 11	41/60	39/60	25/60	24/60	17/60	13/60	17/60	——
* 12	19/30	13/30	16/30	7/30	6/30	7/30	11/30	——
* 13	38/61	36/61	30/61	12/61	14/61	12/61	6/61	——
* 14	32/51	27/51	14/51	5/51	6/51	6/51	10/51	——
* 15	32/48	29/48	24/48	5/48	13/48	10/48	7/48	——
16	——	——	14/29	4/29	13/29	6/29	9/29	——
17	——	——	19/39	8/39	24/39	16/39	9/39	——
* 18	——	——	5/40	10/40	8/40	9/40	9/40	——

* Set had panel exhibit

Table 4

The Number of Frames with One or More Errors
Divided by the Number of Completed Frames
(Second Presentation)

Set	S #3	S #4	S #5	S #6	S #7
1	——	13/54	——	3/54	3/54
2	——	——	——	2/30	2/30
* 3	8/29	12/29	3/29	2/29	3/29
* 4	6/29	14/29	5/29	2/29	2/29
(1st 1/2) 5	8/29	10/29	5/29	2/29	1/29
(2nd 1/2) 5	9/30	15/30	5/30	2/30	1/30
* 6	4/29	6/29	2/29	3/29	3/29
7	3/30	4/30	1/30	4/30	3/30
* 8	5/31	6/31	5/31	1/31	4/31
9	3/29	9/29	2/29	6/29	5/29
10	14/32	8/32	2/32	7/32	3/32
* 11	15/60	18/60	7/60	8/60	6/60
* 12	4/30	8/30	4/30	5/30	8/30
* 13	11/61	23/61	9/61	6/61	5/61
* 14	10/51	17/51	6/51	6/51	7/51
* 15	4/48	6/48	4/48	7/48	8/48
16	3/29	6/29	0/29	3/29	3/29
17	——	19/39	4/39	5/39	6/39

* Set had panel exhibit

Reliability of Latency Measures

Table 5 presents product-moment correlation coefficients based on pairs of read-and-answer values on the first and second presentation of each lesson, for Sets 4-17 inclusive, for the 5 Ss providing such data. S #3 did not have Set 17, the first review test in this program. Degrees of freedom are two less than the number of frames in the second column from the left. Again, asterisks are used to indicate sets with exhibits. There was less interest in the comparison of means since this issue of less overall time on second presentation had been considered previously in the writer's 1961 study. Correlation-coefficient values are occasionally rather high, but there is nothing in the analysis which accounts for the extreme variation among these. None of the highest values in Table 5 seem to be related to extremely low values of Table 6, to be considered next. However, relevant data are limited.

Amount of Material and Read-and-Answer Time

Table 6 has as cell entries product-moment correlation coefficients based on two degrees of freedom less than the number of frames, second column from the left. The rows of the main body of this table are sets of material from The Analysis of Behavior. Note that Set 5 is typically divided into a first and second half in most analyses. Each of the columns contains such coefficients for a different student subject, S. These 8 Ss are identified by a number which both identifies them and gives their rank order with respect to total performance in terms of ratios of errors divided by opportunities for error. S #1 made the greatest number of errors (in terms of most frequent ordinal position with respect to other Ss for each lesson, or set).

Amount of material in each frame was indexed primarily in terms of the number of words. The time, or latency value, for each frame was the time from the button press which advanced the question material into view of S to the press which removed the question and presented the answer.

The correlation coefficients in Table 6 are based on a first presentation of the material, and Table 7 presents coefficients based on a second presentation. In Table 7 less data are available, and the arrangement differs slightly from that of Table 6.

In both tables the correlation coefficients are almost all positive in sign, and some are sizable. The range in coefficient values tends to be greater for those Ss who made the greatest number of errors.

Before considering other analyses related to these tables some background and explanation will be given.

Examined critically, the notion of error rate as sometimes used to describe a lesson, or even a program, seems extremely incomplete. Some of the obvious questions one could ask are as follows: Is a low error rate the result of overprompting? What is the distribution of errors? What is the range on the most-missed or least-missed item?, etc. One type of analysis was concerned with the use of latency in relation to amount of material as a means of adding to the kind of description which seems to

Table 5

Correlation of First and Second Presentation
Read-and-Answer Values

Set	Frames	*S* #3	*S* #4	*S* #5	*S* #6	*S* #7
* 4	29	.40	.63	.66	.76	.56
(1st 1/2) 5	29	.48	.48	.57	.27	.36
(2nd 1/2) 5	30	.80	.63	.58	.29	.43
* 6	29	.53	.80	.69	.68	.70
7	30	.40	.21	-.03	.44	.19
* 8	31	.45	.64	.31	.45	.36
9	29	.41	.58	.38	.26	.43
10	32	.54	.66	.48	.09	.53
* 11	60	.09	.55	.30	.14	.60
* 12	30	.73	.78	.83	.05	.43
* 13	61	.61	.77	.34	.62	.45
* 14	51	.60	.37	.61	.25	.49
* 15	48	.57	.71	.33	.90	.88
16	29	.19	.56	.52	.18	.19
17	39	——	.25	.05	.63	.61

* Set had panel exhibit

Table 6

First Presentation Correlation of Read-and-Answer Times on *The Analysis of Behavior* with Amount of Material

Set	Frames	*S* #1	*S* #2	*S* #3	*S* #4	*S* #5	*S* #6	*S* #7	*S* #8
1	54	.36	.59	.53**	.49	.62**	.45	.56	.17
2	30	.50	.59	.60	.65	.55	.51	.55	.42
* 3	29	.21	.39	.69	.51	.56	.27	.32	.36
4	29	.17	.35	.08	.69	.16	.45	.59	.16
(1st 1/2) 5	29	—	.35	.39	.51	.33	.60	.40	.38
(2nd 1/2) 5	30	.46	.34	.41	.51	.14	.37	.37	.39
* 6	29	.82	.75	.69	.62	.35	.85	.50	.55
7	30	.43	-.01	.19	.05	-.11	.49	.30	.11
* 8	31	.46	.49	.16	.55	.38	.19	.41	.14
.									
.									
.									
17				.37	.43	.37	.54	.61	

* Set had panel exhibit

** Based on first 29 frames

Table 7

Second Presentation Correlation of Read-and-Answer Times on *The Analysis of Behavior* with Amount of Material

Set	Frames	S #3	S #4	S #5	S #6	S #7
* 3	29	-.04	.66	.33	.55	.52
* 4	29	.35	.42	.14	.72	.38
(1st 1/2) 5	29	.21	.54	.50	.43	.56
(2nd 1/2) 5	30	.45	.41	.42	.42	.54
* 6	29	.48	.70	.52	.62	.69
7	30	.31	.39	.15	.32	.21
* 8	31	.14	.42	.24	.31	.54
9.	29	.38	.18	.23	.51	.39
10	32	.51	.37	.50	.29	.36

* Set had panel exhibit

be needed but inadequately provided for by the current notion of error rate.

Some easy frames would seem to provide for little more than a reading task; other frames may pose problems. The time spent on the reading task would seem to be largely a function of amount of material. One finding of the earlier latency study was a product-moment correlation coefficient value of 0.83 for reading time in relation to amount of material in the case of one 66-item programed remedial-English lesson. Correlation coefficients were rarely much greater than 0.20 on the more difficult end-of-lesson review tests, which would seem to require more thinking.

Amount of material was assessed in terms of number of letters, blanks, symbols, lines, arrows, etc. These were simply counted as one unit each. Since a letter count was much more time consuming than a word count, the relationship between the two measures was considered early in the present study. The material of Set 2 of The Analysis of Behavior was assessed both ways. The correlation coefficient (product-moment, as will be true for most coefficients of the present study) was 0.90 for this 30-item lesson, and suggested sufficient commonality between the two methods to justify using the word-count method, rather than the letter-count method.

The relationship between amount of material and time spent reading and answering was examined extensively. A high correlation was assumed to suggest that the student's task was essentially a matter of careful reading. The converse was not assumed, however; e. g., a low correlation was not interpreted as an index of the extent to which the program required an element of "thinking" along with reading. One complication is the possibility that difficult, but brief, items would lead to routine, complete rereading (rather than scanning and selective rereading combined with thinking) and favor a correlation between time and moderately difficult material. However, the heterogeneity of difficulty and length of items in most of the units studied should minimize this. (Another complication will be considered later in relation to the data which seem relevant to it.) Perhaps a more serious complication is that the easier material may not require careful reading. In most material, redundancy is sufficient to render some portions noncritical. However, in the context of programed learning, heavy prompts, as well as redundancy, may favor selective attending to portions of material.

Early in the study, three Ss were asked separately to read aloud a lesson which they had already completed, Set 1 of The Analysis of Behavior. E monitored their oral reading from another room, and the time spent reading the lesson without answering was recorded automatically. The degree of linear correlation between this reading time and amount of material is not as great as was encountered in the previous study by pooling the silent reading of Ss who read in order to answer the questions. However, there is a substantial relationship for the oral reading as indicated by the following correlation coefficients which are each based on 52 degrees of freedom: S #3, r = 0.73; S #4, r = 0.46; and S #5, r = 0.59.

Between one-third and one-half of the correlation coefficients based on silent-reading-and-answering times in the context of normal use of the program fall in the range of coefficient values based on oral reading. In fact, several are even larger.

Taking together the findings for silent reading-and-answering and oral reading, it is evident that there is considerable common variation in the amount of material and the time spent, even though the program under study (*The Analysis of Behavior*) evolved from early revision on a disk-type presentation device and continues to be presented in a format which greatly limits the amount of material which is presented in a single frame.

Of greatest interest, however, is what accounts for the variation among Ss and among program sets in values of these correlation coefficients in the present study. It was assumed possible that amount of material would be less of a determiner as material became more difficult, so that the values of the correlation coefficients of Tables 6 and 7 would be lower on the sets which were more frequently associated with higher error rates.

Errors were used to index difficulty four ways. First, the number of frames with at least one error was divided by total frames in each set for each S and these values ranked. Differences in average rank could then index differences among sets. Second, the same procedure was carried out using only the half of the Ss who made the fewest errors on the program. Third, the per cent of frames in each set missed by more than one-half the Ss was used. A fourth index was based on combining the three measures.

In addition to the coefficients of Table 6, which were considered in relation to these various measures, the average time spent on each frame for all Ss who provided such data was obtained. The correlation based on pooled latency data with amount of material was then considered. Correlation coefficients were then considered in relation to error assessments of difficulty, according to the four kinds of values, discussed above.

In most instances the linear correlation between read-and-answer times (or latency) and amount of material tended to be lower for sets of greater difficulty (as indicated by error data), but correlation coefficients were rarely statistically significant. When the correlation of pooled latency data with amount of material was considered separately for those sets for which there was an exhibit, a Spearman rank R = -0.52 was obtained based on N = 8. For sets without exhibits, R = - 0.31 and N = 9. For sets with and without exhibits combined, R = -0.37 and N = 17.

When the correlation coefficients in Table 6 were paired with the errors/error-opportunities values for all Ss for each set, and these pairs ranked for S #1 to S #8, the rank order correlation coefficients were the following, respectively: 0.81, -0.10, -0.20, 0.45, 0.05, 0.57, 0.13, and 0.13 (N = 9 except for S #1 where N = 8). The degree of linear association between the amount of material, as indexed in the present study, and the time spent reading and answering questions posed by this material does not seem to vary in relation to variables considered in this analysis.

One reason may be suggested by R = 0.81 for S #1. This S, for whom the task seemed far too difficult, seemed to spend very little time on frames which were difficult for her to answer. This points up one more reason why the correlation between amount of material and time spent on this material may not change systematically with difficulty. Another complication is that sets of a well-revised program may not differ greatly in difficulty, defined by error data, from set to set.

In concluding this type of analysis, the correlation coefficients were used as x-members of pairs with S's own errors/error-opportunities ratio as y-members. It was regarded possible that changes in r-values would covary more regularly with changes in S's own indicator of difficulty than with changes in indicator based on the pooled error data of all Ss. For S #1 to S #8, respectively, rank order correlation coefficients were as follows, respectively: -0.12, -0.02, -0.15, 0.62, -0.61, -0.62, 0.18, and 0.14. For Ss #s 1, 3, 4, 5, and 8, N=8. N =9 for all other Ss. Again, the correlation between amount of material and latency does not change systematically with change in difficulty even when S's own error data are used to index this difficulty.

A final analysis, dealing with this topic in the case of The Analysis of Behavior program, involved the selecting out of frames on the basis of how frequently each was missed. These were selected from Sets 4 - 8 inclusive, sets on which there were data from 8 Ss for most frames. Separate product-moment correlation coefficients were calculated for read-and-answer values in relation to amount of material for those frames never missed, and for frames missed by 1, 2, 3, 4, 5, and 6 and 7 Ss combined. None of the frames in this material were missed by 8 Ss. The respective correlation coefficients are given with the associated degrees-of-freedom value as follows: 0.88, 7; 0.77, 26; 0.55, 25; 0.81, 25; 0.23, 30; 0.83, 13; and 0.29, 6. The association of time and amount of material is generaly fairly strong, but variation in the degree of this linear association does not seem related to changes in the difficulty of material as indexed by error data.

These data provide one of the best demonstrations of the relationship between latency and error data. For each frame that had been missed by 0, 1, 2, 3, 4, 5, and 6 and 7 Ss combined there was a total of the latency values provided by the 8 Ss. The totals based on the sum of the totals for 8 Ss for each frame were obtained separately for 0, 1, 2, 3, etc., error frames. The number of frames on which each mean was based can be obtained by adding two to the degree-of-freedom values given previously. From the zero-error frame category to the one which includes frames missed by 6 or 7 of the 8 Ss, the means of latency totals were, respectively, as follows: 133, 172, 178, 203, 251, 280, and 236. The rank order correlation coefficient of these values with errors is R = 0.86, N = 7, and P<0.05, one-tailed test. Low latency values are to be found on low error frames.

In order to obtain the value of each latency on a per-person-per-frame basis these values must be divided by 8, e.g., 133/8.

Note that there is an increase in read-and-answer values from 0-error frames to 5-error frames, inclusive, and that these average latencies are more than twice as great for 5-error frames as for 0-error frames. The 6-and-7-error mean poses a question: is it less than the mean for the 5-error frames because at this level of difficulty some Ss soon begin to give up trying to answer, or is it merely a sampling error since it is based on only 8 latency totals?

This type of analysis was repeated. The only difference was that instead of using read-and-answer values, such values were each divided by the amount-of-material value and the resulting values which reflect time per unit of material used instead. Starting with zero-error frames, the averages for these values were as follows: 5.48, 6.73, 8.03, 7.36, 8.83, 9.48, and 9.21. The agreement between these ranks and the rank of number-of-error categories is even higher than before, although there are several one-rank changes in order: $R = 0.93$, $N = 7$, $P < 0.01$, one-tailed test.

Set 4, the second half of Set 5, and Sets 6, 7, and 8 of The Analysis of Behavior were also considered in another analysis. Product-moment correlation coefficients were obtained separately for each of these by using the sum of the latencies for 8 Ss as x-values and the number of errors associated with each frame for y-values. For Set 4, the second one-half of 5, and through 8 inclusive, in the same order, those coefficient values and associated degrees of freedom are as follows: 0.57, 27; 0.42, 28; 0.43, 27; 0.63, 28; and 0.35, 29.

This same procedure was then carried out with a substitution of time-per-unit-of-material measure, as obtained in the second of the two analyses just described involving means, for the sum for 8 Ss of times to read and answer. Again, starting with Set 4, the coefficients and degrees of freedom were the following: 0.61, 27; 0.20, 28; 0.18, 27; 0.33, 28; and 0.44, 29.

In comparing these values with those just obtained without considering the amount of material, it appears that two of the five are slightly larger but that the three which became smaller are markedly reduced.

In considering the relationship between latency and error data further, an attempt was made to add a refinement in the case of error data. In previous, less systematic study, a suggestion offered by Jack Michael[1] was considered and found worthy of further study. He suggested that, as a first approximation, errors be weighted according to source by arbitrarily dividing the number 100 by the number of errors each S had made on any given unit of material. The more errors a given S made, the smaller would be the value of each error as such values are used for revising the sequences.

---

[1] Michael, J. L. Personal Communication

After arriving at a per-error value for each S, these were to be assigned to each frame on which each S had made an error; then a total-error factor was arrived at by summing these for each frame as an indicator of frame difficulty. According to this method, smaller total values would be assigned to a frame missed by many Ss, each of whom had made a great number of errors, than to a frame missed by fewer Ss if each of them had made few errors.

Separate product-moment correlation coefficients were obtained for Set 4, the second half of Set 5, and Sets 6, 7, and 8 using weighted-error values as x-members of each pair and latencies as y-members (the sums of read-and-answer times for the 8 Ss). As before, there were pairs of values for each frame in the set. Starting with Set 4 the correlation coefficients and degrees of freedom are as follows: .49, 27; .36, 28; .43, 27; .28, 28; and .29, 29.

In summary, the relationships between latency data and error data have been examined in several ways. (Presentation of information in tables has included in left-hand columns data for less appropriate Ss for The Analysis of Behavior program, and on the right, data for students of increasing appropriateness. The tendency for longer latencies to be associated with more errors has characterized all analyses. First, median values of read-and-answer times for those frames in each set missed by Ss have been greater than the median for frames in the same lesson answered correctly, a statistically significant finding. Second, the less-often-missed frames are associated with lower read-and-answer latency values, and there is an orderly association between error and latency values. An increase in one is accompanied by an increase in the other. When expressed in terms of a rank-order correlation coefficient, the association is stronger if, instead of latency, one uses latency per unit of material, indexed largely in terms of the number of words. The relationship between latency and error holds with a reduced range of r-values when errors are weighted so that errors made by the better students count more in the analysis. However, differences among sets of frames with respect to different ways of measuring the variables of errors, latency, or even amount of material, no doubt, are yet to be explained.

In general, it does appear promising to develop an index of difficulty which is based upon errors, latencies, and amount of material for the purpose of program revision. However, an attempt failed to refine one component of such a measure by analysis of the correlation between latency and amount of material.

Incidental additional data have been provided, often mainly to provide the reader a means of looking into some matters of interest to him but overlooked by E. The details of error data on a second presentation are an example. The extreme variation in reliability of read-and-answer values for the first as compared to the second presentation, is another.

Independent Factual Items

Another approach to the study of relations among amount of material, error data, and latency was afforded by the isolated-facts materials devised by Goldbeck (Goldbeck and Campbell, 1962). As initially developed, his materials consisted of a "program" (of 35 frames covering 35 topics) which was presented in one of three versions to different groups of Ss. He gave a test later which was the same for all groups and designed to be equally favorable to all versions of the material. In the initial Goldbeck-Campbell investigation, during training one group got an easy, heavily prompted version; a second group received a program of moderate difficulty; and a third received a difficult-item version. Replication of each of these conditions occurred in terms of three response modes in the Goldbeck and Campbell study.

By selecting randomly the level of difficulty of the item covering each topic, a set of material was developed for use in the present study. In contrast with the procedure in the Goldbeck-Campbell study, each S had some easy, some medium, and some difficult items. The same test developed by Goldbeck was used, and the order of topics during training and test was the same.

On one-half of the items for each of the three difficulty levels S did not get to see the program answer after writing his response to an item until after a delay of 10 seconds. For each item one half the Ss encountered such delay, the other half no delay. Programing of delay was semi-automatic. The only relevance this delay should have for the results considered for the moment is, perhaps, to increase the variance of the latency measures. The results of this delay procedure are not part of the present study.

Table 8 presents the average latencies (read-and-answer times) based on 12, 12, and 11 easy, medium, and difficult items, respectively, for 7 Ss. Note that these values are all lowest for the more difficult items. Further, the longest average latencies for some Ss are on easy items, for others, on ~~difficult~~ medium items.

Table 8

Average Latencies for Isolated-Fact Items

First Presentation

	S#1	S#2	S#3	S#4	S#5	S#6	S#7
12 Easy Items Average:	33.17	31.83	36.83	22.75	28.67	22.75	19.58
12 Medium Items Average:	32.83	28.00	35.25	26.83	33.58	17.58	20.00
11 Difficult Items Average:	25.55	17.18	30.36	20.73	21.00	17.36	15.45

One purpose in using these materials was that the greater planned variance in difficulty could point up possible complications in the relation of error to latency. For example, it seemed possible that Ss would simply identify some frames as being too difficult. The only read-and-answer time which would accrue to such frames would be the time it took S to recognize a frame he would not be likely to answer.

The finding of lower average latencies on the most difficult items is counter to the general finding of an increase in latency with an increase in difficulty and would seem to favor an interpretation that S did not spend much time on items he regarded as impossible to answer correctly.

The results of Table 8 point up the importance of taking amount of material into account. Easy items were made easy through redundancy as well as by having prompts. When the per cent of total time is considered for each category and this is divided by the percent of total material of which the items in that category are constituted, it is evident that the most time per unit of material is spent on the most difficult item and the least on the easy ones: Easy items, 37.4% of time/47.8% material; Medium items, 36.9% time/36.8% material; Difficult items 25.7% time/15.4% material.

Table 9 pertains to the correlation of read-and-answer time with amount of material for these items at the three levels of difficulty during the two training presentations and for test items prepared for by training on items of that level of difficulty. The numbers identifying Ss are the same as assigned previously and, thus, reflect their overall performance with respect to errors on The Analysis of Behavior material. Note the increase in correlation-coefficient values for the first presentation of Easy items starting with the S who made the most such errors until S#4, whereupon there is an orderly decrease. Again for Easy items during test, there is orderliness in the decrease from S to S starting with S#7 to the left.

Table 9

Correlation of Read-and-Answer Time with Amount of Material for Three Difficulty Levels for a First Presentation of Training Followed by Test. A Second Training Presentation Came Last.

		S#1	S#2	S#3	S#4	S#5	S#6	S#7	All Ss
Easy (n=12)	1st Presentation	.37	.50	.63	.73	.60	.49	.12	.76
	Test	.27	.05	.11	.36	.41	.56	.58*	
	2nd Presentation	.75	.76	.24	.65	.12	--	--	
Medium (n=12)	1st Presentation	.40	.68	.48	.43	.36	-.26	.51	.58
	Test	-.06	.16	.32	-.31	.05	-.20	.48*	
	2nd Presentation	.52	.50	.62	.43	.32	--	--	
Difficult (n=11)	1st Presentation	.18	.23	.32	.50	.46	-.01	.27	.50
	Test	.16	.33	.31	.16	.23	.41	.41*	
	2nd Presentation	.61	-.17	.89	.73	.30	--	--	
All Levels Difficulty Combined (n=35)	1st Presentation	.52	.76	.42	.36	.52	.24	.29	.69
	Test	-.02	-.11	.18	.11	.22	.43	.41*	
	2nd Presentation	.51	.66	.69	.58	.46	--	--	

*S had 4 presentations before test.

For the twelve items of medium difficulty, the correlation coefficients, reflecting covariation of read-and-answer time with amount of material, differ less when first and second presentations are considered than comparable coefficients based on the same number of easy items. They also differ less than those based on the eleven difficult items.

In the case of Difficult items a progressive increase in correlation-coefficient value from S#1 to S#4, followed by a decrease starting with S#5, is noticeable, similar to that for these same Ss in the case of Easy items.

It is interesting that the correlation of latency with amount of material based on the pooling of read-and-answer times for all Ss is highest for Easy, next highest for Medium and lowest for Difficult items on the first presentation, the presentation for which there is complete data. It is also interesting that the rank order of TEST, EASY correlation coefficients agrees so well with the rank of Ss based on errors/ error-opportunities values on The Analysis of Behavior: R = -.89, N = 7.

Incidentally, there is, no doubt, greater interest on the part of many readers in the relationships among errors, latencies, amounts of material, etc. on test performance than performance on the program *per se*. It was intended that half the *S*s on *The Analysis of Behavior* take the test, before repeating earlier material, and half not, but instead go back to Set 1 from Set 16. Poor timing, attrition of *S*s, etc., resulted in limited data. However, these did suggest an interesting possibility. It seemed that *S*s missed fewer test questions when these were prepared for on sets which had higher error rates. However, *S*s did seem to miss more test questions when they missed the *most relevant* program item leading up to the test. While this is merely speculation so far as the data of *The Analysis of Behavior* is concerned, since there was very little data and the relevance of program items to test items is difficult to establish, it is a possibility for further study. Perhaps students can learn more by answering easy teaching items which are surrounded by relatively more difficult irrelevant items. Perhaps the motivational and attentional benefits afforded by the difficult items put *S* at an advantage. This advantage would be cancelled by the errors he makes on the difficult items, and *S* may do less well on the test items for which he prepared, using training items which were too difficult. However, in the case of easy items, there is nothing to cancel the benefit afforded by surrounding difficult items. The empirical question, thus, is whether or not students learn more from easy items in a context of easy items or from easy items in a context of difficult ones; and, if there is a difference, what are some of the relevant parameters.

In the isolated-fact materials, it was possible to relate teaching items to test items. Also, three levels of difficulty had already been established for all items. It was interesting that there were fewer errors on test items for which *S*s had prepared by taking easy training items. In the present study, overt, composed answering was the response mode. In the Goldbeck-Campbell study, *S*s who were trained with all items of medium difficulty, and who used this same answer mode, did best on the test. A major difference between the present study and the Goldbeck-Campbell study is that each *S* in the present study dealt with some items at each of the three levels of difficulty, whereas in the Goldbeck-Campbell study, each *S* worked on items at only one difficulty level. In the present study a teaching machine was used; Goldbeck and Campbell used booklets.

Three analyses of variance were performed which used the number of errors as scores in terms of each of the three difficulty levels. The first training presentation was followed by the test about one week later. A day or so after the test, *S* again repeated training as carried out during the first presentation. (The purpose of posttest training was, in part, to continue to study the effects of a 10-second delay in feedback--a peripheral aspect of the present study.) The plan to repeat the test after the second training presentation was not carried out; some *S*s were lost before this phase of study could be completed.

In the treatment by subjects analyses of variance, and *t*-tests which followed, the error values for the difficulty training items (and test items trained for by difficult items) were corrected by a factor of 12/11 since there was one fewer of these items than items of both the easy- and medium-difficulty levels.

Table 10

Summary of the Three Treatments X Subjects Analyses of Variance for Two Training Presentations and Test (which Followed the First Training)

Source	df	1st Training		Test		2nd Training	
		MS	F	MS	F	MS	F
Difficulty Level	2	92.79	53.64*	9.60	5.65***	39.43	11.88**
Subjects	6	12.96		44.37		21.35	
Interaction	12	1.73		1.70		3.32	

$^{*}P < .001$ $^{**}P < .005$ $^{***}P < .025$

Fewest errors were made on Easy items during both training presentations and on the corresponding test items. The means, corrected as explained previously, for errors on the first training presentation for Easy, Medium, and Difficult items were, respectively, as follows: 3.29, 8.86, and 10.13. A *t*-test for Easy and Medium means indicated that $t = 7.92$ and $P < .001$. Similarly, for the same ordering, means on the test were 5.43, 7.71, and 7.01. Here the *t*-test was for Easy and Difficult means since this time the value of the Difficult-item mean was slightly closer to that of the Easy-item mean than the Medium-difficulty mean. The *t*-value was 2.27 and $P < .05$. For a second presentation which followed the test Easy-, Medium-, and Difficult-item means were, respectively, as follows: 1.43, 4.57, and 6.08. A *t*-test, which again involved the mean number of errors on Easy- and Medium-difficulty items, indicated $t = 3.22$ and $P < .01$. In all of these *t*- tests, the error term was that obtained in the analysis of variance reported above, therefore, df = 12.

The analyses of variance and *t*-tests were based on procedures suggested by Lindquist, 1953.

Latitude

Table 11 is based on data obtained in a single session in which one presentation of a draft set of geography frames, dealing with latitude and developed by the writer, was followed immediately by a second presentation of the same material. Moreover, *S*s were working under special incentive conditions in which they could earn a fifty-cent bonus for perfect performance, or some fractional bonus if their performance was near perfect. In addition, it was suggested that how they compared with other *S*s was now of interest. Both aspects of these instructions were intended to provide for unusually valid and reliable measures. Immediate repetition of the lesson was also intended to favor reliability. The two rank order correlation coefficients which reflect first and second presentation reliability for the two kinds of correlation-value measures are based on N = 7. R = 0.714, or greater, is required for statistical significance at the 5 per cent level for N = 7, one-tailed test. Again, there seems to be nothing included in the present study which accounts for the variation in the coefficients of the first two rows of Table 11.

Table 11

Latitude: A 30-Frame, First-draft Set
First 5 Rows of Cell Entries are Product-Moment Correlation Coefficients

		S#1	S#2	S#3	S#4	S#5	S#6	S#7	
Time with Material, Two presentations	1st 2nd	.13 .18	.55 .54	.19 .37	.47 .41	.38 .29	.30 .28	.45 .45	R = 0.86
Time with Errors, Two presentations	1st 2nd	.33 .35	.18 .12	.26 .48	.33 .38	.34 .44	.25 .27	.25 .30	R = 0.77
Time on 1st with time on 2nd presentation		.73	.47	.13	.48	.73	.51	.70	
Number of Errors Two presentations	1st 2nd	23 13	26 16	22 10	18 4	9 1	5 2	7 1	

Earth-Sun Relationships

A first-draft version of a program dealing with the earth-sun relationships was of interest both as an example of an unrevised program and as a program using many multiple-choice items as well as items which required a sentence or so for an answer.

An inspection of some of the test items which followed this program, and for which data were first analyzed, suggested that latency data would not have the usual relation to error data. They would be useful in pointing up items which were missed because the student misread them, however. For example, one item was rather lengthy, was answered incorrectly, but very quickly, and dealt with the beginning date of spring from the viewpoint of the southern hemisphere.

However, the program portion did provide a sufficient number of frames to conduct the same kind of analysis as was performed with The Analysis of Behavior materials. Items were selected and grouped on the basis of the number of student errors associated with each item. There were 14 zero error items, 21 one-error items, 13 two-error, 10 three-error, 14 four-error 17 five-error and 3 six-error-items, cr frames. Based on sums for all 6 Ss, the average latencies on these, respectively, were as follows: 230, 273, 260, 346, 339, 353, and 298. The order of these average sums of latencies is in statistically significant agreement with the order based on number of errors associated with the frames on which the average was based: $R = 0.714+$, $N = 7$, $P = 0.05$, one-tailed. When these latencies, or read-and-answer times, were each divided by the amount of material in the frame on which they were based, the following values were obtained, listed in the same order as previously: 4.45, 6.43, 5.90, 7.91, 7.74, 8.56, and 9.30. There is greater agreement between the order of these values and error-rate ranks than for time values without consideration of amount of material: $R = 0.93$, $N = 7$, $P < 0.01$.

Two conclusions are suggested. The relation found previously between latency and error data in the case of The Analysis of Behavior program holds in the case of a program which is in a first-draft stage and which contains both many multiple-choice items and sentence-length composed-answer items. The relation between latency and error data is found to be stronger when latency-per-unit-of-material values are used to partial out variance resulting from differences in amount of material. (It should also be recalled that, under the extreme conditions which held in the study of isolated-fact items, failure to consider amount of material may obscure the general relation.)

Symmetry and Rotation Program

This program typically required answers with more parts and much more time for S to answer. It was extremely difficult for most Ss. Insufficient data were obtained to justify lengthy analysis.

Perhaps the most interesting findings were the casual observations related to use of this material: Some Ss, when given a choice of program following use of this material, protested, saying, "Anything but more of this kind of material!" However, what was interesting was that Ss who were previously satisfied with The Analysis of Behavior now wanted the program dealing with verbal problems in algebra. When questioned in more detail, some indicated that while they found the psychology program more interesting, it was harder than algebra. They made more errors. After making many errors it seemed especially important to get back to a learning task which was the least challenging. Normally, material of greater interest compensated for occasioning more errors.

Other Observations from the Study

One student, S#3, commented that the most interesting session had been when she had the delay following her answer. Delays seemed to her at first to follow items which she had answered correctly. Delay seemed to heighten her interest in answers when she was uncertain of their correctness.

S#7, in participating in the feedback-delay study, described above in relation to the isolated factual materials, reported unusually effective learning conditions. E chose this S for the first session, and S got the benefit of a faulty circuit design. Instead of the answer being first brought into S's view after a 10-second delay, S got a glimpse of the answer just as the projection lamp cooled at the beginning of the 10-second delay in feedback. With the question and correct answer to mull over for 10 seconds, S regarded the task as easy to master and the best learning conditions he had encountered during the study.

Other observations concern E's various kinds of incidental information concerning each S in relation to overall error rate on The Analysis of Behavior. The S who had the lowest error rate was a college junior, the next lowest, a college freshman with a higher, overall grade average than the next S , who was also a college freshman, etc. In brief, there was good agreement between relative standing in terms of error rate and level and quality of educational background. As the tables of this section have suggested, Ss also tended to maintain the same relative standing on various learning tasks. This is not surprising in view of the gross differences among Ss with respect to most kinds of data.

The analyses considered so far have touched on a variety of issues. One major issue has been the relationship between latency and error rate. The general and statistically significant finding has been that greater latency values go with a higher error rate, lower latency values, a lower error rate. If there is considerable variation in amount of material, this relationship may be obscured. Consequently, the relation between time-per-unit-of-material values and error rate may hold with greater generality. The errors of poorly motivated students, or those poorly qualified in terms of the prerequisites assumed for the program, may not be as closely related to characteristics of the program as the errors of the most ideal student users. Weighting errors according to student source may aid in demonstrating commonality between latency and error. However, in the present study the relation between weighted-error values and latency merely seemed to vary less for different lessons.

The attempt to relate changes in the degree of linear association between amount of material, indexed largely in terms of words, and latencies (read-and-answer times) to the difficulty of material, in terms of errors, cannot be considered very encouraging.

It was found that frames with the lower error rate favored fewer errors on the test covering the material taught by these frames. It was suggested that difficulty may favor learning, and error oppose learning. One may have his cake and eat it too by embedding easy items in the context of items which are more difficult. Important to this suggestion was the finding by Goldbeck and Campbell that students who trained on only moderately difficult items did best on the test. In the present study, students had items at three levels of difficulty during training.

One S commented favorably on how interesting the training task was because of the 10-second delay between completion of an item and opportunity for scoring. Another S commented that giving an answer via a quick flash on the screen immediately and then again after 10 seconds (during which time S mulled over the question, his answer, and the correct answer) was an ideal condition for learning. This S supported his comment by a perfect score on the test. However, this could also be the result of the extra two training trials or the result of this S's unusual capability. Nevertheless, it is possible that some parameters of delay, especially between trials, or frames, could benefit both learning and thinking about what is learned.

Selected Frames

Frames, or items, were selected according to error, latency, and amount of material criteria from The Analysis of Behavior program, Sets 1 through 18, inclusive. The amount of data among these sets differed, and the conclusions arrived at inductively should stand only as tentative.

The per cent of time spent on each frame of each set was obtained by considering each of the 19 units (first one-half and second one-half of 5 were considered separately) as being made up of 100 per cent. The value assigned a given frame was the per cent of the total time spent on the set which was spent on that particular frame. The per cent of the total amount of material, as defined previously, largely in terms of number of words, in each set was similarly obtained for each frame. A ratio of per cent time divided by per cent material was then obtained for each frame.

In a sample of 673 frames, 22% were less than 20 words in length. More than 45% of the frames which had a % read-and-answer time/% amount-of-material ratio of 1.2, or greater, were this brief.

A value of this ratio of greater than 2 was obtained for frame 25 of Set 2 (2-25) which reads, "In this unconditioned hand-withdrawal reflex, heat is the ________ ________." Apparently, the two students or so on whom these data were based were forced to stop and think in order to relate a newly-acquired technical term to an example. The brevity of this item, however, suggests that, in more absolute terms, this item posed a minor thinking requirement. Frame 29 of Set 5 (5-29) is an example of a frame which demands very little of S: "Many reflexes occur together in a(n) ________ state." For data based on 6 Ss the ratio was 1.36, but this frame required the least time and contained the least material of any frame in the first one-half of Set 5. Frame 2-25 would seem to be superior to frame 5-29 in the challenge it affords the student.

Frame 4-21 asks the student to read about an experiment, gives two columns of data, and asks S to give technical terms for each. None of the 7 Ss who provide data missed either part of this frame. In several instances, the use of panel material increased the time spent on frames without increasing the error rate, e.g., 3-2, 3-6, 3-7, and 3-21.

Frame 5-21 was answered correctly by each of the 7 Ss. However, 2.7% of the time spent on the 29 frames of this half of the set was spent on this frame which contained 4.6% of the material. Frame 5-21 is essentially a copy frame. Frame 5-27 is similar in each of these respects, but proportionately more time is required. The copy requirement is less verbatim; S is supplied "increased" and expected to write "rise(increase)." Another frame which demands proportionately little of S, in terms of the ratio of time to amount of material, as well as in an absolute sense, is frame 5-39 which calls for a simple "yes" or "no" answer. Again, none of the Ss answered incorrectly.

In the context of the program these frames may have done their job well. However, it seems likely that they could have continued to serve the intended function if they had been slightly more demanding. In terms of the present analysis, it would seem that the verbatim copy frame, 5-21, could most stand greater task loading. Another, rather extreme example, is frame 6-9: "Salivation is the action of the salivary ________." Students spent less than 1% of their time on this frame which contained somewhat more than 1% of the material of this set. The absolute demand of frame 7-1 is greater, but the material portion of the time/material ratio is much greater, suggesting that there is little challenge by this item which requires S to state that "The behavior of a hungry animal can be 'rewarded' with ______." In frames 7-28 and 7-29 none of the 7 Ss failed to answer correctly; but, the ratio of time to material comes closer to 1, with both the relative and absolute amount of time greater for frame 7-29 than 7-28. Both of these frames may demand more of the student by using asterisks which require the student to fill in as many words as seem necessary rather than blanks which suggest the number of words intended.

Frames 8-10 and 8-11 were both answered correctly by 8 of the 8 Ss. Frame 8-11 required both relatively and absolutely more time and more time relative to amount of material. In frame 8-10 the major behavior-determining element seems to be ". . . _______ of response." In frame 8-11, S gives two responses which constitute a step toward his discriminating between reinforcement and reinforcer. More would seem to be gained from revising frame 8-10, as by removing the two words after the blank and replacing the blank with asterisks, than by further task loading of 8-11.

Examples of especially thought-provoking frames would seem to be frames 3-25 and 14-28, although these are considered in relation to the data of only 4 and 5 Ss, respectively. Frame 3-25 seems to take S by surprise by asking after, and in the midst of, many details of an experiment, in essence, "What was it all about?" It reads, " In Pavlov's famous experiments on conditioning, ________ was the unconditioned response."

In frame 14-28, there is reference to an experiment which provides an example, if a critical feature is noted, of a kind of behavior which is defined in Set 14: "Since in the experiment the relation between receipt of food and ritualistic behavior is accidental, we say the behavior is ________."

One purpose of this section is to point up some of the unevenness left in certainly one of the best programs of its kind. Looked at from the standpoint of how they could profitably be made more, rather than less, difficult, several examples of minimal challenge to student users have been cited. The worst items would seem to be copy frames, (semi-copy, perhaps), and frames with strong prompts, especially near the blank. Simple questions which require "yes" or "no" answers may require little more of the student, but there may be greater conceptual benefit from these frames. The use of panel materials may lead the S to spend proportionately more time per unit of program material and engage him in a task which enables him eventually to deal with relatively complex material without error. Good examples of questions which Ss could answer correctly, but which they could not answer immediately, were frames involving the relation of definition to example and critical differences between similar technical terms. Especially good frames may be those which get S to respond to recently-considered details in terms of some basic aspect.

New Latency-Defined Characteristics

Some of the benefits from this project are yet to be realized in promising starting points which are now little more than speculation.

One approach to introducing greater challenge is to introduce special prompts which indicate to S the relative difficulty of different items. Informed that 70% of all students missed a given item, S might give it the attention and caution required for success.

Another innovation which would make use of both latency and error data is largely new only in the way it would be implemented. Test items could be graded in difficulty and arranged from easy to difficult. In addition, however, exceptionally difficult items which deal with trivial and peripheral matters could be seeded in this series.

In the start toward an experimental program, the set dealing with latitude, the intention was to orient the material toward a pool of transfer items, items requiring complex application of what the program taught. Accordingly, any time a problem could be conceived which related to the basic subject matter, but which could not be solved by a program user, a revision would thereby be suggested. Such an approach would be complementary to both error and latency analyses and related revisions.

In response to the brief article by Gerlach, Baker, and Schutz (1964), which points out that students sometimes can answer questions along the way only to become unable to answer these questions by the end of the program, latency analyses should be useful in identifying those correctly-answered items which are most in need of review from time to time.

The following relates to the matter of errorless testing. By using observations of long latencies on relatively simple test items which are responded to correctly as predictive of errors on the same kind of subject matter, if greater complexity were introduced, the appropriate educational sequence for any particular student could be provided. The particular value of latency suggestive of a high error rate on more complex items would have to be arrived at experimentally. However, the correlations observed between latency and error rate suggest that such a calibration could be feasible. A gain in overall student learning efficiency, and, if the implications of Terrace's work apply, a gain in dependability of newly-acquired behavioral capabilities should result.

Summary

This chapter considers the performances of individual students. A small, rather heterogeneous sample of students, with respect to the appropriateness of background for some of the learning tasks, has provided the data of this chapter. Most of the students participated for many sessions, over a period of two or three months or so. They also worked on several kinds of learning tasks.

A basic assumption has been that new methods are needed which use more aspects of student performances as guides to program revision. Current practices which rely too exclusively on error data lead to undetected portions of the program which are too easy. Student data should also be obtained which point up portions which should be revised to be more difficult. It is suggested that learning may be more efficient as a result of such revision and that the student may be more inclined to think at greater depth about what is learned.

Latency analysis, as a complement to error analysis, is suggested as a method. Latency information can be obtained on every frame, errors on only those, eventually few, frames which are answered incorrectly. Latency tells more about frames which are too easy. Some index of the amount of material is also likely to be helpful. There are suggestions in various portions of the present study that latency per unit of material is of greater general usefulness, e.g., the isolated-fact materials.

Recommendations of another investigator concerning the potential value of latency data were noted. However, the present study interprets latency as more of a predictor of errors than as another term for reaction time. Undoubtedly there is broad commonality between latency during learning and latency after learning has taken place.

Analytical study dealt with student errors, latency, and amount of material. Each of these was indexed in more than one way, and variant forms of operational definitions were considered. Some index for describing a set of frames, or a program, in terms of difficulty seems to be needed. It would permit one to compare different programs in terms of relative difficulty, perhaps one correlate of the amount of thinking required of the student. It would also permit a programer to see how much change there was in difficulty from one revised version to the next. The use of only error rate for this purpose was regarded as unsatisfactory.

A fruitless investigation of one component of an alternative index was pursued at length. An effort was made to use latency in relation to amount of material for this purpose, as a complement of error rate. The correlation between latency and amount of material would suggest that each portion of the program was of some influence on student behavior as the coefficient, which reflects degree of this association, increased in value. However, student response to the program seems far too complex to be described in such simple terms. This conclusion grew out of many detailed attempts to find some relation between changes in r-values and various indicators of difficulty based on error data.

This may have related, in part, to unexplained changes in latency value from a first to a second presentation. Examples of this kind of reliability were given. Sometimes there would be a high correlation between first and second presentation values. For the same person, or for the same lesson, sometimes there would also be very little correlation.

Despite such suggestions of the unreliability of latency, strong and statistically significant relationships were found of an increase in latency values going with increases in error rates. This was demonstrated in terms of the data of individual students. For all students the median time spent on frames which were missed was more often greater than the median time spent on frames answered correctly. Also, when frames of The Analysis of Behavior program which had been completed by 8 Ss were selected on the basis of the number of errors associated with each, the rank order based on the number of errors was in high and statistically significant agreement with the rank of these frames in terms of read-and-answer latency values. A similar finding was obtained using a first-draft program which had many multiple-choice as well as one-sentence answers. In both cases, $R > .9$ when latency was dealt with as a residual after allowance was made for amount of material. (Latency values without such "correction" were also statistically significant, but less so.)

One further refinement was to use weighted values and assign greater weight to errors which were made by those students who made the fewest errors. The correlation between values based on weighting errors and latency was similar to the correlation based on non-weighted values, simply the number of errors. However, r-values, based on the same 8 Ss but different sets of material, tended to vary less from set to set. (The number of pairs of values was the number of frames in each set since each frame was the unit with which pooled latencies and pooled errors were associated.)

The effects of errors on learning were considered in two ways. First, a casual observation suggested that after S had made many errors he preferred an easier-than-usual program. Second, a formal study involving items at three levels of difficulty--items associated with three degrees of error rate--indicated that less-often-missed items taught better. An important feature of this study was that teaching items were clearly related to test items. A previous study using the same test and the same overt, composed-answer mode found items of medium difficulty to be superior for Ss who trained exclusively on items of medium difficulty. This would tend to suggest that the test was not biased in favoring easy items. An important difference between the previous study, which was performed by Goldbeck and Campbell, 1962, and the present one was that in their study different groups of Ss used items which were all at one of the three levels of difficulty. It is possible that easy items teach better when surrounded by more difficult items. One complication in this interpretation, however, is that the control afforded in the present study by the teaching machine may have increased S's difficulty in dealing with items at the medium-difficulty level, changing the actual level upward.

Illustrative use was made of latency data and information concerning amount of material in the case of The Analysis of Behavior program. Several frames were considered which were answered correctly by all of the 8 Ss in the present study. Some of these required as little as one per cent of the total time S spent on a lesson. The excessive simplicity of such frames is even more apparent when such frames contain a per cent of material of the lesson which is greater than the per cent of time they require of students. Example frames are part of this argument.

On such a basis, sample frames were noted which all Ss answered correctly which did require relatively more time. Several of these involved material (panels or exhibits) outside the program. Another example required that S respond to a basic aspect of the lesson likely to be obscured by details. The present study, by using a number of marginally appropriate Ss, was poorly suited to pointing out those frames of greatest challenge and least hazard of error. These benefits of latency analyses could best be demonstrated by being part of the evolutionary process during program revision.

Such a start toward an experimental program led to only a first draft, which was used under conditions most likely to show unusually high first- and second-presentation reliability of latency and latency-derived measures.

The intended approach was to evolve a brief program, with many of the characteristics of existing ones, which was revised in terms of failure to prepare students for transfer items, as well as test items more directly related to the program. An arrangement of problem items in graded difficulty, from easy to difficult, was contemplated. Greater difficulty was to be attained by letting the student be informed, or cautioned, of what per cent of a test group had missed each especially difficult item. Rather than allowing the student to become excessively confused, and later hampered in learning a difficult portion of the subject matter, it was suggested that the most important aspects be taught with easy items surrounded by difficult items dealing with less crucial topics.

Errorless testing could be part of this process. Materials could be experimentally calibrated, as improved measures of latency and error are developed, so that long latencies on rather easy items could be used to warn of errors if the student were confronted with more complex, related items. The student could later deal with these materials after completing additional material which increased his chances of success.

ABSTRACT

Purpose

The purpose of the project was to utilize programed instruction in order to (a) stimulate increased thinking activity on the part of the student, and (b) gain additional information concerning the nature of thought processes and how they may be stimulated.

To achieve the purpose, the characteristics of linear programs were analyzed experimentally in order to search for descriptive dimensions of such programs which enhance active responding by the student. In addition, new program formats were developed, and some of these were used in a classroom experiment concerning methods of enhancing ability to solve problems.

A rational analysis of conventional programing techniques suggested that there should be variations within programs in the style of sequences intended differentially to establish new learning, to provide review to enhance retention, and to give more practice to the student in organizing and consolidating his new learnings.

Procedure

A preliminary phase of work consisted of developing samples of new program formats and exercises based upon a rationale for improving the degree of active responding and organizing activities of the student. The practice of interspersing such new exercises among the sequences in a linear program was recommended on the basis of consideration of the task, criterion, and individual differences variables which, taken together, argue for a flexible rather than a rigid program style.

In a classroom experiment, one new format was administered under various procedural conditions intended to appraise the extent to which the requirement to ask questions and to utilize hints in the practice of solving problems would differ from additional study of the basic materials of the lesson, in terms of increased ability to solve other problems pertaining to the same substantive content.

In a series of laboratory studies with individual students utilizing linear programs, analyses were made of the characteristics of frames which appear to encourage the degree of active responding needed for effective learning.

Results and Conclusions

Groups which were required to solve practice problems following study of the basic principles did not perform any better on a criterion test consisting of other problem-solving tasks than did groups which spent additional time studying the basic lesson. This held true whether or not the groups required to solve practice problems were also required to ask questions about the practice problems or to utilize a series of hints

provided to them. As in other studies, time spent in study appeared more closely related to problem-solving criterion performance than did variations in how the study was directed. In this case, knowledge of principles, coupled with the student's own ability, appeared to be the main factors in successful problem solution.

Analysis of records of individual students taking linear programs under laboratory conditions indicated that use of response latency measures in conjunction with program errors would provide a better basis for program revision than does the present standard procedure of revising on the basis of error data alone. One reason for this is that latency measures are obtained on all program frames, not merely upon frames missed by the student. Latency measures also are more useful in explicitly pin-pointing the faulty teaching frames which result in poor criterion performance. Since, in an instrumented condition, latency can be observed before any overt response is made, error could be avoided by providing a prompt after a long latency (difficult item), and short latencies could indicate where more difficult frames could safely be introduced. In short, latency techniques can help implement the recommendation that frames be cued close to response threshold but not much above.

Some new formats were developed which were not used in formal experimentation, due to time limitations. These program formats and the rationales for them are reported in order to encourage future research on these promising variations from current programing procedures. A possibility suggested by one analysis is that key materials could be emphasized successfully by (a) insuring that key responses are sufficiently prompted to lead to a correct response, and (b) surrounding key frames with less important frames of greater difficulty.

While the complexity of the problem of thinking leaves many gaps to be filled between the analyses conducted with linear programs and the experimentation conducted with drastic departures from conventional programs and procedures, it is believed that continuation of these two very different appearing lines of investigation could result in eventual merging at a common meeting point.

References

Amsel, A. "Error responses and reinforcement schedules in self-instructional devices." In A. A. Lumsdaine and R. Glaser (Eds.) *Teaching Machines and Programmed Learning: A Source Book*. Washington, D. C.: National Education Association, 1960. 506-516.

Anderson, R. C. and Anderson, R. M. "Transfer of originality training." *Journal of Educational Psychology*, 1963, *54*, 300-304.

Angell, D., and Lumsdaine, A. A. "Prompted and unprompted trials versus prompted trials only in paired-associate learning." *Student Response in Programmed Instruction*. (Edited by A. A. Lumsdaine.) Washington, D. C.: National Academy of Sciences--National Research Council, 1961. Chapter 24, 389-398.

Briggs, Leslie J., and Angell, David. "Programed Instruction in Science and Mathematics." *Review of Educational Research*, June, 1964.

Briggs, Leslie J., and Hamilton, Nancy Russell. "Meaningful learning and retention: Practice and feedback variables." *Review of Educational Research*, December, 1964 (in press).

Briggs, Leslie J., and others. "Experimental Results Regarding Form of Response, Size of Step, and Individual Differences in Automated Programs." *Programmed Learning and Computer-Based Instruction*. (Edited by John E. Coulson.) New York: John Wiley and Sons, 1962. Pp. 86-98.

Brooks, Lloyd O. *Response latency in programmed learning: Latency related to error rate*. Unpublished doctoral dissertation, University of Houston, 1961 (now available through the services of University Microfilms, Ann Arbor, Michigan).

Campbell, Vincent N. "Bypassing as a way of adapting self-instruction programs to individual differences." *Journal of Educational Psychology*, 1963, *54*, 337-345.

Crutchfield, R. S., and Covington, M. V. *Facilitation of creative thinking and problem solving in school children*. Paper read at meeting of AAAS in Cleveland, December 1963.

Ferster, Charles B., and Sapon, Stanley M. "An Application of Recent Developments in Psychology to the Teaching of German." In A. A. Lumsdaine and R. Glaser (Eds.) *Teaching Machines and Programmed Learning: A Source Book*. Washington, D. C.: National Education Association, 1960. 173-185.

Gagne, Robert M. "The Acquisition of Knowledge." *Psychological Review* 69: 355-365; July 1962.

Gagne, Robert M. "Problem Solving." *Categories of Human Learning*. (Edited by Arthur W. Melton.) New York: Academic Press, 1964. Pp. 293-317.

Gagne, Robert M., and Brown, Larry T. "Some Factors in the Programing of Conceptual Learning." *Journal of Experimental Psychology* 62: 313-321; October 1961.

Gagne, Robert M., and Dick, W. "Measures in a self-instructional program." Psychological Reports, 1962, 10, 131-146.

Gerlach, Vernon S., Baker, Robert L., and Schutz, Richard E. Measurement procedures in programed instruction. Programed Instruction, February 1964, III, No. 5, p. 4.

Goldbeck, Robert A., and Briggs, Leslie J. "An Analysis of Response Mode and Feedback Factors in Automated Instruction." Programmed Learning. (Edited by Wendell I. Smith and J. William Moore.) New York: D. Van Nostrand Company, 1962. Chapter 11, pp. 166-191.

Goldbeck, Robert A., and Campbell, Vincent N. "The Effects of Response Mode and Response Difficulty on Programed Learning." Journal of Educational Psychology 53: 110-118; June 1962.

Goldbeck, Robert A., and others. Integrating Programed Instruction with Conventional Classroom Teaching. San Mateo, Calif.: American Institute for Research, December 1962. 28 pp. (Mimeo.)

Hamilton, Nancy Russell. Identification and Quantification of the Confirmation Variable in Research. Paper read at a meeting at the American Institute for Research, Washington, D. C., 21 March, 1964. (Mimeo., 21 pp.)

Haselrud, G. M., and Meyers, S. "The transfer value of given and individually derived principles." Journal of Educational Psychology, 1958, 49, 293-298.

Hershberger, Wayne A., and Terry, Donald F. Delay of Self-Testing in Three Types of Programed Text. Palo Alto, Calif.: American Institute for Research, May, 1964, (AIR-C28-5/64-TR). Mimeo.: 17 pp.

Holland, James G., and Skinner, B. F. The Analysis of Behavior. New York: McGraw-Hill, 1961. 337 pp.

Kersh, Bert Y. The attainment of complex objectives: Teaching discovery processes and increasing motivational effects. Paper read at an Invitational Working Conference on Programed Instruction, National Educational Association, Washington, D. C., March 24-26, 1963.

Kersh, B. Y., and Wittrock, M. C. "Learning by discovery: An interpretation of recent research." Journal of Teacher Education, 1962, 13, 461-468.

Lambert, W. E. "Measurement of the linguistic dominance of bilinguals." Journal of Abnormal and Social Psychology, 1955, 50, 197-200.

Levin, Gerald R., and Baker, Bruce L. "Item Scrambling in a Self-Instructional Program." Journal of Educational Psychology 54: 138-143; June 1963.

Lindquist, E. F. *Design and Analysis of Experiments in Psychology and Education*. Cambridge, Mass.: Houghton Mifflin Co., 1953. 393 pp.

Lumsdaine, A. A. "Some Conclusions Concerning Student Response and a Science of Instruction." *Student Response in Programmed Instruction.* (Edited by A. A. Lumsdaine.) Washington, D. C.: National Academy of Sciences--National Research Council, 1961. Chapter 31, pp. 471-500.

Melton, Arthur W., editor. *Categories of Human Learning*. New York: Academic Press, 1964. 356 pp.

Robinson, Francis P. *Effective Study*. New York: Harper and Row, 1941.

Russell, D. H. "Higher mental processes." In Harris, C. W. (Ed.), *Encyclopedia of Educational Research*. New York: Macmillan, 1960.

Siegel, P. S. "Concept formation: Response-time considerations." *Psychological Reports*, 1964, 435-442.

Skinner, B. F. *Science and Human Behavior*. New York: Macmillan Co., 1953.

Skinner, B. F. "Final Report to the Fund for the Advancement of Education." August 15, 1958.

Skinner, B. F. "Teaching Machines." *Scientific American*, November 1961, 205, No. 5, 90-102.

Suppes, P. "Modern learning theory and the elementary-school curriculum." *American Educational Research Journal*, 1964, 79-93.

Terrace, H. S. "Discrimination learning with and without 'errors.'" *Journal of the Experimental Analysis of Behavior*, 1963, 6, 1-27. (a)

Terrace, H. S. "Errorless transfer of a discrimination across two continua." *Journal of the Experimental Analysis of Behavior*, 1963, 6, 223-232. (b)

Terrace, H. S. "Errorless discrimination learning in the pigeon: Effects of chlorpromazine and imipramine." *Science*, April 1963, 140, No. 3564, 318-319. (c)

Wittrock, M. C. Verbal stimuli in concept formation: Learning by discovery. *Journal of Educational Psychology*, 1963, *54*, 183-190.